



## **Vocal Bodies**

### **Performing Paralinguistic Stereotypes and Multivocalities in Art and Digital Media**

Hasse Jørgensen, Stina Marie

*Publication date:*  
2020

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY-NC-ND](#)

*Citation for published version (APA):*  
Hasse Jørgensen, S. M. (2020). *Vocal Bodies: Performing Paralinguistic Stereotypes and Multivocalities in Art and Digital Media*. Det Humanistiske Fakultet, Københavns Universitet.

# Vocal Bodies

## Performing Paralinguistic Stereotypes and Multivocalities in Art and Digital Media

*Stina Marie Hasse Jørgensen*



PhD Dissertation  
Department of Arts and Cultural Studies  
University of Copenhagen

Stina Marie Hasse Jørgensen  
Vocal Bodies: Performing Paralinguistic Stereotypes and Multivocalities in Art and Digital Media  
Department of Arts and Cultural Studies, University of Copenhagen

2020

Main Supervisor: Gunhild Borggreen, Associate Professor, Department of Arts and Cultural Studies, University of Copenhagen

Co-supervisors: Holger Schulze, Full Professor in Musicology, Department of Arts and Cultural Studies, University of Copenhagen. Jakob Kreutzfeldt, Head of Festival: Struer Tracks – Urban Sound Art Festival

Proofreading: Macon Holt

Front Page Illustrations:

*Collective Performative Reading*, NLH Space 2016. Photo Credit: Frida Gregersen  
*The [multi'vocal] voice recorder box*, Catch Art and Technology Center, 2018. Photo Credit: Nina Cecilie Højholdt



## Acknowledgements

No project this large would ever be possible without a considerable amount of encouragement and support. First thank you to Department of Arts and Cultural Studies, University of Copenhagen for granting me the scholarship which allowed me to embark on this project.

Unlimited thanks to my fellow collectivists and friends who participated in the two collective vocal projects developed during the PhD process. You helped shape the research conducted throughout this project – there would not be a dissertation if it was not for you; Frederik Tollund Juutilainen, Alice Emily Baird, Mads Pelt and Nina Cecilie Højholdt from the *[multi'vocal]* collective, and Jane Jin Kaisen with whom I share the collective *For More Than One Voice*. Jane, I also hold a deep gratitude for your generous reading and comments on my writing, and for all the inspirational and supportive talks we had throughout the many years of this PhD.

I want to direct a great many thanks to my main supervisor Gunhild Borggreen. Thank you for your attentiveness to my writing, which made me improve on all fronts. This has been a long process and your encouragement, advice on my research and writing has been invaluable. Thank you also to Holger Schulze, my co-supervisor, for pointing me towards the intriguing phenomenon of Hatsune Miku, and for interesting conversations on smart voice assistants and speakers during the last couple of years. I also had the pleasure of being co-supervised by Jakob Kreutzfeldt in the start of my PhD project. He was instrumental in directing me towards the sound design of robots, for which I am very grateful.

Boundless appreciations to my colleagues at the research network ROCA (Robot Culture and Aesthetics), and fellow PhD students at the Department of Arts and Cultural Studies – especially a huge shout out to my sisters in the extended work group *Daughters of 68*; Katrine, Cecilie, Sabrina, Sofie, Tobias, Heidi, Inger, Anne, Rasmus, Kristine, Emilie, Pernille, Grishma, Tina, Majken, Mette and Solveig. Thank you for all the inspiration, good conversations, energy and support. Without you I would not have managed to take pleasure in writing this dissertation.

Endless gratitude to my dear discussion partners and readers Katrine, Alice, Frederik, Tobias and Jane, my grandmother and my grandfather who have been supportive readers and conversation partners in discussions of writing, art, and politics over the years, and not least my mom, who have supportively, tirelessly and generously been reading countless versions of what has become this PhD dissertation. Also, a great thanks to Jonas Fritsch for giving crucial feedback in the final sprint of this writing.

Thanks to Center for Computer Music at Columbia University for accepting me as visiting scholar at a crucial moment in the process of my practice-based research. Without this stay I would not have met Alice Emily Baird and formed the collective ongoing project *[multi'vocal]*, which has become a crucial part of my practice-led research. Thank you especially to artist Douglas Repetto, one of the most amazing teachers I have ever had, for sincere interest, continuous support and for having opened my world to sound technology and meaningful explorative approaches to sound art ten years ago.

Thanks to the IT University of Copenhagen and the Digital Design Department for giving me a beautiful shared room with a view the last couple of years. I truly appreciate how the institution creates space for experimentation in practice and theory, and fosters a spirit of collaboration on projects, across faculty and students in its labs and hallways. I am looking forward to continue the collaborations.

I also want to express huge appreciation for the many students at University of Copenhagen, Lund University and the IT University of Copenhagen, who in different ways have all contributed to the development of the analysis and projects that this PhD dissertation presents. Thanks to all the artists who have participated in interviews and fruitful correspondences. I especially want to extend warm thanks to Nanna Lysholt Hansen, Niels Pugholm, Nalini Malani, and Jane Jin Kaisen.

On a personal note it has been crucial to my wellbeing, creativity and sanity throughout the years to share my life with you Nana, Ida and Anine. I am grateful for all the dinners, parties and vacations we shared through the last fifteen years in 'firkløveren'. May, my old

school friend – thank you for opening my thoughts to playful imagination and for being the most amazing friend through highs and lows. And Martez, filosnack, thanks for all the discussions about music and the meaning of life – to be continued. And also a great thanks to all my other friends who have made life good and fun during the PhD years.

A very special gratitude and indebtedness to my family; my Timrilt family, my father and the Jørgensen-clan, my uncle Morten and Christian, bedste Thomas, and especially my grandmother for always being there for me and being a role model for how to write with passion, and to my mom for her immense love, encouragement, support and inspiration in life and work. Massive gratitude must also go to my in-laws, grandparents Bodil and Ulrik, and uncle Terkel – you are the best and most loving family I could ever dream of being part of. You give so much love, peace and care.

Most of all, I thank you, my most amazing, brainy, fun and passionate partner Steffen, for your enormous love. It means the world to me to share my life and love – not least to our daughter – with you.

This thesis has taken many years to write. Years where I have lost my dearly loved grandfather Werner and given birth to my loved-beyond-words daughter Sonja Marie – who has brought new meaning to my life in a completely wonderful, loving and insistent way every early morning, day and night. The dissertation is dedicated to Werner and tiger-Sonja for having been and for being in the world, making it a better place with all the love they share.





# Table of Contents

|  |            |
|--|------------|
| <b>Introduction.....</b>   | <b>13</b>  |
| The Question.....  | 14         |
| The Relevance.....   | 16         |
| The Contribution.....  | 19         |
| The Structure of the Dissertation.....   | 20         |
| <b>Paralinguistic Technologies.....</b>  | <b>26</b>  |
| Vocal Auditive Meaning.....  | 29         |
| Performing Paralinguistic Bodies.....  | 31         |
| <b>Listening to and through Voices.....</b>  | <b>40</b>  |
| Interpretive, Embodied, Situated and Relational Listening.....   | 41         |
| Performative Listening.....  | 45         |
| Practice-based and Practice-led Research.....  | 47         |
| Collective Practice-led Research.....  | 52         |
| The Political Context of Collective Work.....  | 57         |
| Practice-led Research and Collective Feminist Work.....  | 59         |
| <b>The Articles.....</b>   | <b>65</b>  |
| ROBOT & THE END: A Comparative Critical Reading of the<br>Staging of Synthesized Voices in Digital Media Performances (2018). 68       |            |
| The Politics of Voices in Contemporary Art (2019).....   | 83         |
| Vocalizing Dystopian and Utopian Impulses (in press, 2020).....  | 94         |
| <b>The Collective Vocal Projects.....</b>  | <b>118</b> |
| Collective Performative Reading (in press, 2020).....  | 119        |
| [multi'vocal]: Reflections on Engaging Everyday People in the<br>Development of a Collective Non-Binary Synthesized Voice" (2018). 138 |            |
| <b>Conclusion.....</b>   | <b>152</b> |
| Paralinguistic Stereotypes.....  | 153        |
| Paralinguistic Multivocalities.....  | 155        |
| Summing Up and Pointing Forward.....   | 160        |
| <b>References.....</b>   | <b>169</b> |
| <b>Abstract.....</b>   | <b>176</b> |
| <b>Resumé.....</b>   | <b>177</b> |
| <b>Appendix.....</b>   | <b>178</b> |
| The Politics of Voices in Contemporary Art (translation).....  | 179        |
| Technological Fantasies of Nao – Remarks about Alterity<br>Relations (2017).....   | 190        |



The voice does not move away from my body, but rather it carries it forward - the voice *stretches* me; it drags me along, as a body bound to politics and poetics, its accents and dialects, its grammars, as well as its handicaps.

(Brandon LaBelle: *Lexicon of the Mouth*)

**You can cut out the cards dispersed  
in the dissertation and place them  
on a floor or somewhere else**



# Introduction

“The voice is undoubtedly one of the most noble and most intensively researched issues in the field of sound studies,” writes professor of Sound Studies and Musicology, Holger Schulze (2015). The cultural anthropologist Amanda Weidman states that the voice is absolutely central to cultural, social and political life (2014). Weidman particularly emphasizes the dual nature of voice as both a sonic material phenomenon and voice as a culturally elaborated metaphor where “... the realms of the cultural and sociopolitical link to the level of the individual, a site where shared discourses and values, affect, and aesthetics are made manifest in and contested through embodied practice” (2014, p. 38). Voice is both a matter of expression and of being heard, and connects deeply to feelings of intimacy, identity, bodily practices and imaginaries. The voice has received new attention from the theoretical and artistic scenes since the early 2000s. During which time there has been a vast production of books, articles, and artworks that address the voice through its philosophical, linguistic, sonic, and bodily aspects, as sound scholar and artist Norie Neumark remarks in her introduction to the book, *VOICE* (2010). My PhD project is part of this development, as it presents research on the performative bodily aspects voice in contemporary art and digital media performances. In this dissertation, it is not my intention to scrutinize every aspect of the voice, or to present an overview of all the different theoretical concerns and theorizations of the voice. It is rather my intention to engage with the performative, auditive, technological and bodily aspects of voice. Specifically, I focus on how vocal bodies are performed, technologically mediated, modified and synthezized in art and digital media in connection to the *paralinguistic-body*. I have developed the paralinguistic-body as a term, in order to capture the many different relations between paralinguistic performance, identifications and notions of bodies - including the imaginary bodies. Paralinguistic means in short, alongside linguistics and refers to the non-lexical elements of communication in and around speech.

The studies on paralinguistics performances of vocal bodies conducted in this dissertation are situated within sound studies, connecting to fields of audio engineering and computer science. It is an important new field of study as the vocal bodies, whether real or imaginary, are part of shaping how we come to understand ourselves and others through a material relationality of vocal expressions.

## The Question

In this PhD project I am interested in investigating what role technologies play in the paralinguistic performance of vocal bodies as stereotypes or as multivocalities.

In the very start of this PhD project I investigated the sound design of social robots<sup>1</sup> and found that the categorizations of the computational paralinguistic traits and states – such as pitch, timbre, pace and other vocal features beyond speech that can be experienced as e.g. accent, gender, age of synthetic voices designed for social robots and smart assistants – operated with e.g. stereotype accents, ages and binary gender stereotypes of male or female (see Jørgensen, 2018). This tickled my curiosity, and I started to wonder since synthetic voices are generated and performed through machines and are therefore not limited to a single vocal identity, why then do synthesized voices have only one gender, one age, and one accent? Could there be more participatory, inclusive and imaginative alternatives? Could synthetic voices represent more than one voice, identity and body?



Experiments with the social robot Nao at Medical Museion, 2015.

Audio engineers- and computational scientists defines the paralinguistic as the study of speech beyond the communicated message within computational paralinguistic studies (Baird et al, 2017; Baird et al, 2018; Schuller and Batliner, 2014). In this PhD project, however, the paralinguistic is understood more broadly as the bodily, performative and

---

<sup>1</sup> See Jørgensen and Tafdrup (2017) for more information about this investigation.

sonic qualities and imaginaries of vocal expression. This understanding draws on the artist and sound studies scholar Brandon LaBelle's conception of 'cultures of the paralinguistic' (2014). And it is relation to this field of study that I would like to direct attention to the question that has informed my research:

In which ways do technological manipulation, mediation and synthesis of vocal expressions create listening experiences and imaginaries that perform, negotiate and complicate paralinguistic-bodies in art and digital media?

My inquiry begins as an exploratory practice as well as an analysis of arts practices, which make audible paralinguistic performances of vocal bodies. These range from performances of individual to collective bodies, creating listening experiences of paralinguistic stereotypes and paralinguistic multivocalities.

Etymologically 'multivocality' or 'multivocal' is a combination of the words 'multi' and 'vocal' and literally means 'many-voiced,' and is defined as signifying many things: of manifold meaning. As ethnographer Robert Mizzi writes multivocality can provide a space for plural voices to be heard (2010). Mizzi draws on the philosopher and literary scholar Mikhail Bakhtin's work on language where he uses the notion of "multi-voicedness" to demonstrate that a "plural consciousness" can be useful when writing autoethnography. Mizzi writes that Bakhtin critiqued the use of a single voice as being a "monologic" discourse, and instead promotes a "dialogic" discourse that emerges through the exercise of writing a novel or an essay to locate, organize and include multiple identities, desires and voices within the human subject (Bakhtin, 1981). In this PhD project, my focus is different from Mizzi (and Bakhtin) in that I am here not focusing on multivocality of speech as such, but on the multivocality of the paralinguistic materiality of vocal performance. Professor of communication Patrice M. Buzzanell writes that multivocality destabilizes the singular and unified in, for example, research processes of analysis and representation, which opens up space for alternative interpretations, multiple identities, roles and standpoints (2017). Also, Lisa Jansen and Michael Westphal writes about multivocality as something where multiple identities and roles are created in speech or singing (2017). Here Jansen and Westphal – like Buzzanell and Mizzi – focus on the multivocality of



speech and language, and not the multivocality of the paralinguistic materiality of vocal performance as I do in this PhD project.

In this dissertation I use multivocality to suggest that meaning and representations of performative vocal expressions are not fixed in a unified way. Instead, meaning and representation of the performative paralinguistics of voice are constructed in relation to the context and to others. In this way, paralinguistic multivocality is a concept that questions the expressions of voice as markers of fixed stereotyped bodies, performing paralinguistic-bodies as always partial, divided, transgressing and becoming in relation to other humans and non-humans, as shared affective experiences and imaginaries. In this dissertation, I present several different ways in which paralinguistic multivocality is performed in art and digital media, as auditive bodies of transnational and transhistorical trauma, emigration, consumption, gender oppression, and othering in the neoliberal consumer society.

## **The Relevance**

The research question stated above have guided my PhD project – ever since my first experiments with the vocal performances of social robots (2014-2016). This question is important to ask and investigate right now, because we are today witnessing an increasing interest in the design of voices and vocal identificatory bodies.<sup>2</sup> In the age of ubiquitous computing, voice user interfaces have the ability to complement, and even replace the graphical user interface. The sonic qualities of voice have strong impact on the acceptability of the systems in which they are embedded, and quality issues of vocal appearance are even subject to public discussions – journalists have diagnosed it as the ongoing “voice wars” (Yoffie et al, 2018). Companies have invested on a scale of millions of dollars to develop designs for human sounding voices implemented in transportation systems, vending machines, conversational interfaces, smart assistants such as Siri by Apple, Alexa

---

<sup>2</sup> For a very long-time machines did not speak with human sounding voices, and it is only recently that companies have been able to design human sounding synthetic voices. Before the mid-90s most synthesized voices were created with signal processing techniques, modelling the human vocal tract such as the formant synthesizer *Votrax Type-n-Talk* chip (Gagnon, 1978). This changed with the development of the method of concatenative synthesis, which trained models on human voice recordings (Hunt and Black, 1996). Today machine learning methodologies are used. Here, large amounts of tagged data are used to “teach” machines the performances of vocal expressions based on voice samples. The voice samples are usually collected from a single voice actor. For example, the voice samples recorded from the voice actress Susan Bennett trained the vocal expressions of Apple’s smart assistant Siri. Siri as well as Amazon’s Echo and Google Home today provide options of male and female voices. Recent developments such as Deep Neural Network based systems are used to further model the emotional expressiveness of the voices (Watts et al, 2016; Zen, Senior and Schuster, 2013).

by Amazon, Cortana by IBM and Google Home by Google, and in the voices of social robot assistants such as Softbank Robotics' Nao or Pepper. These voice-based assistants have created new conditions for interactions between people and computers leading to new emotional relations, desires, imaginaries and identifications with the technologies. Designer Bert Brautigam describes how the synthetic voice of today is progressively designed with anthropomorphic traits, interacting with everyday people through gendered human-sounding speech synthesis, and often promoted as something human subjects can identify with and relate to:

It speaks to us like a person. We call it names. It throws in a joke every once in a while. We attribute emotions and feelings to our voice assistant. It defines itself as female or male. The sound of a voice assistant imitates human sound and intonation. (Brautigam, 2017)

This development of making synthetic voices sound more and more human-like, has been debated by many scholars (Baird et al, 2017; Baird et al, 2018; Bickmore and Cassell, 2005; Harris, 2005; Juutilainen, 2019; Jørgensen et al, 2018; Lee, Nass and Brave, 2000; Männistö-Funk and Sihvonen, 2018; Nass, Moon and Green, 1997; Nass et al, 2001; Nass and Brave, 2005; Phan, 2017; Tamagawa et al, 2011; Tompkins, 2011; Yen and Nass, 2010). This debate is concerned with the need for interfaces to imitate human behavior, and asks if synthesized voices should adopt anthropomorphic traits, such as gender, dialect, and portrayal of emotion. Digital Media scholar Frances Dyson writes in her book *The Tone of our Times* that the synthesized voice has become a 'who' to which we relate through the computing and quantification of the tone of the voice:

The quantification of vocal tone, the synthesizing of the voice, and the development of human-computer interfaces based on human-machine conversation, create agents that provide the "who" in this post-theological era. (Dyson, 2014, p. 69)

The vocal tone, which can be experienced when synthesized voices are designed to imitate human voices through designs of paralinguistic features such as age, gender, accent and emotional expressions are important because, as anthropologist Jenny R.

Lawy reminds us, “[s]ignifiers such as age, gender, race...and many other factors serve to filter and discern who speaks clearly and powerfully, and who is ignored” (2017, p. 198). The paralinguistic signifiers of the synthesized voice matter as to ‘who’ is reiterated as a powerful and clearly heard vocal body. Media scholar Thao Phan writes in her article, “The Materiality of the Digital and the Gendered Voice of Siri” (2017), that the quantification of vocal tone as either male or female is reiterating gender binary norms of visible subjects.

In spring 2018 the CEO of Google, Sundar Pichai, impressed his audience with a presentation of the newest development of the company’s human sounding and heavy American accented young male and female synthesized voices. The demonstration of the synthesized voices showed that these kinds of voices could make a reservation of a table at a restaurant or an appointment for a haircut over the phone, with the person at the other end not recognizing that they were speaking with a machine (*DR Nyheder*, 2018). It is off course technologically impressive to make machines sound like humans to the extent that they trick listeners into thinking that the machines have an identity and subjectivity – but why design the machines to have voices representing normative gender, accent and age? Phan answers that this is because the human sounding voices of smart assistants can make users forget that they are communicating with an algorithm calculating or retrieving information. When the human interlocuters hear the voice of what they believe is an audible intelligible subject, the technological mediation become almost invisible. Phan argues:

A successful natural user interface is, then, one that does not challenge the user; it is one that is invisible to them. Invisibility here means diverting attention away from the act of mediation – it is the illusion of immediacy by which the subject is perfectly seduced by the medium, and in this seduction, indulges in the fantasy that there is no medium at all. (Phan, 2017, p. 28)

Phan presents an important dimension of the paralinguistic design of synthetic voices as something that is used to make listeners to forget that they are interacting with machines, promoting normative cultural understandings of vocal bodies in an attempt to make the interaction with digital media appear seamless. The adherence to these norms can be said

to connect to the valuation of the individual over the collective in neoliberal capitalism (Crawford and Joler, 2018; Dyson, 2014; Jørgensen, Vitting-Seerup and Wallevik, 2017). The growing tendency of designing voices performing stereotyped singular bodies demands critical inspection. However, it also makes it relevant to look into alternative articulations and performances of voices, as well as to explore other expressions and experiences of paralinguistic-bodies. In this way, the computational paralinguistic designs of the performative vocal materiality insert machines into debates on personification and open up for important discussions of alternative vocal bodies and imaginaries that challenge normative vocal representations, which is highly important to discuss and explore today.

## **The Contribution**

With this PhD project, I contribute to the studies of vocal expressions in sound studies, audio engineering and computer science, by investigating how vocal expressions are stressed and transformed through technological mediation, distribution or synthesis. I listen to and through the voice as a paralinguistic-body performing stereotypes or multivocalities. I present different performances of paralinguistic multivocalities that move experiences of vocal bodies away from associations with identity politics to that of structural critiques or imaginaries of alternative worlds and conditions – in an interplay with the listener. I present how this interplay is unfolded through works in contemporary art and digital media performances as presented in three articles included in the thesis, and in the collective practice-led vocal projects *Collective Performative Reading* (2016-) by the collective *For More Than One Voice* and *[multi'vocal]* (2017-) by the collective *[multi'vocal]*, presented as a remediation and as documentation included in the dissertation. The discussion has been developed through an interplay between the theory and practice that has unfolded during the PhD project. The discussion contributes to a more nuanced analysis of how the technological paralinguistics of voices in contemporary art and digital media perform vocal bodies. It also informs new participatory ways of designing voices in digital media. In this way, this PhD thesis contributes to the field of audio engineering and interaction design, in the presentation of an alternative design as it is explored in *[multi'vocal]*.

The vocal practice-led projects furthermore propose a research format that is collective and auditory. This approach also contributes to the new movement of sonic dissemination of research within the field of sound studies, which traditionally has presented research on sound through the written word (Abu Hamdan, 2012; Boss, 2018; Holmboe, 2019; LaBelle, 2014; Lane, 2006; Young, 2015). I use this dissertation to explore an alternative format for presenting research on voice: the written text will be interrupted with quotes from the *Collective Performative Reading* as a remediation consisting of a series of tasks in which the reader is momentarily asked to listen to her own paralinguistic features as something that inform and are informed by the relations to other vocalizing humans or nonhumans. These periodical interruptions will make the reader listen to the performative sonic qualities of the voice, and their many bodily expressions and imaginaries which this thesis is all about. In doing so it proposes an alternative research format to the field of sound studies and the field of computational paralinguistic in audio engineering.

### **The Structure of the Dissertation**

The combined PhD dissertation consists of three chapters, three articles, and two texts documenting or remediating the collective vocal projects. The purpose of the chapters is to articulate a common framework through which the articles as well as the collective research projects can be understood as a coherent project.

The **first chapter** presents the concept of paralinguistic technologies by combining the notion of ‘cultures of the paralinguistic’ from sound studies with computational paralinguistics. This meta-theoretical framework is behind the analysis in a more fragmented form when discussed in the included articles, and in the contributions that present the two vocal projects – but is presented coherently in the first chapter. Here I first present how voices have been discussed as auditive meaning within computational paralinguistics and within sound studies, suggesting a combined field of research concerned with paralinguistic technologies. Then I present how voice has been discussed as something performing the body. Based on these theoretical discussions I develop the concept of paralinguistic-body.

The **second chapter** presents how I listen to and through performative paralinguistic-bodies as a methodological approach. First, I present listening as relational, embodied, situated and performative. I argue that this approach to listening makes way for including the listener in the discussion of performances of vocal bodies. Then I present a methodology of listening through practice-led research and how I have worked with this through collective practices.

**The three articles** with discussions of vocal performances in contemporary art and digital media are included in the combined dissertation under their titles: 1. “ROBOT and THE END: A Comparative Critical Reading of the Staging of Synthesized Voices in Digital Media Performances” (2018), 2. “The Politics of Voices in Contemporary Art” (2019), “and 3. “Vocalizing Dystopian and Utopian Impulses” (in press, 2020). All the three articles are peer-reviewed, one as a journal article and two as contributions to anthologies.

**The collective vocal projects** are presented as remediation and as documentation: 1. “Collective performative reading” (forthcoming, 2020), and 2. “[multi’vocal]: Reflections on Engaging Everyday People in the Development of a Collective Non-Binary Synthesized Voice” (2018). The texts on the collective vocal projects as well as the articles should be considered as documentation of an iterative process which has unfolded throughout this PhD project – as a journey through which the exploration of vocal bodies as paralinguistic performances of stereotypes and multivocalities has gradually evolved into the most important concern across different analysis of contemporary art, digital media and collective practice-led research.

The **concluding chapter** of the dissertation presents a reflection on what vocal paralinguistic performances in art and digital media (as it is presented in the included articles) and the vocal practice-led projects (as it is presented in the included remediation and documentation) can contribute to the field of paralinguistic technologies. Furthermore, I present how the paralinguistic-bodies in the works discussed in the articles and the collective vocal practice-led projects perform paralinguistic stereotypes or multivocalities. Finally, I define and propose possible future research areas in connection to the research which has emerged during the development of the PhD project.



Ciò che comunica è precisamente l'unicità vera, vitale e percepibile di chi la emette. Non si tratta però di una comunicazione chiusa nel circuito fra la propria voce e il proprio orecchio, bensì di un comunicarsi dell'unicità che è, al tempo stesso, una *relazione* a un'altra unicità. Ci vuole almeno un duetto, un chiamarsi e risponderci: ossia una reciproca intenzione di ascolto, già attiva nell'emissione vocale, che rivela e comunica ognuno all'altro.

(Adriana Cavarero, *A più voci: filosofia dell'espressione vocale*)



**You can choose and read out a quote at any time and at your own pace**



## Paralinguistic Technologies

Technology can enable situations and conversations that would not otherwise have taken place. As media and sound theorist Douglas Kahn writes in his book *Noise, Water, Meat: A History of Sound in the Arts* (1999), technology can lead to a transformation of materiality, a resolution or change of meaning in a situation. Technology can also alter, stress and transform vocal expressions. But what role does technology play in the performance and experience of vocal expressions as bodily identities? How are technologically mediated voices altering or challenging experiences of bodies? What bodies are tied to digital or collective voices? These are some of the questions that I have discussed in the articles included in the PhD dissertation, and which will be presented in the conclusion.

In order to ground the discussion, I here present how the voice has been theorized as computational as well as bodily, material, performative, specifically focusing on the paralinguistic dimension of vocal expression as it has been theorized within sound studies as well as in computational paralinguistics. Apart from the two co-written texts on computational paralinguistic (Baird et al, 2017; Baird et al, 2018), the focus on the paralinguistic is not explicitly addressed in the articles and contributions included in the dissertation as this conceptual framework has emerged as my studies evolved. However, this meta-theoretical perspective highlights the commonalities of the different explorations by showing how they all contribute to a shared field of study, generating a new relevant perspective on the discussions of voice within sound studies and computational paralinguistic. In the following, I elaborate on the research done in relation to the specific focus on paralinguistics, wherein I discuss the voice as technological, auditive and performative.

The book *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing* (2014) by Björn Schuller and Anton Batliner is the first book that surveys computational paralinguistics from automatic speech, to speech signal processing and machine learning. The authors write that the field of 'computational paralinguistics', relating to fields such as philology, information technology and audio engineering, started

in the late 1990's with Rosalind W. Picard's studies on emotion and speech, in her seminal book, *affective computing* (1997), which presents studies of affective computational vocal production. Schuller and Batliner write that studies of paralinguistics – meaning 'alongside linguistics' – is concerned with *how* you say something rather than *what* you say and define the computational paralinguistics as a field that is concerned with how affect, emotion, personality and everything else beyond linguistics – the 'para' linguistics – can be expressed by or embedded in human speech and language through computation.

Paralinguistics (without computation) were, at first, an area of study confined in relation to human-human communication. Here the body language, gestures, facial expressions, tone and pitch of voice are all examples of paralinguistic features (Trager, 1958). The linguist David Crystal's definition of paralinguistics later defined a narrower field of study focusing on "the tones of voice" scoping the term paralinguistic as excluding visual communication specifically addressing "vocal factors involved in paralanguage." (1975, p. 163). Crystal explains that the vocal factors are vocal effects lingering "at the edge of language," that is the vocal effects in and around words including laugh, cough, sigh, and yawn etc. which "seem to have an important role to play in the communication of meaning" (ibid.). Crystal sees the paralinguistic phenomena as "*any* meaningfully contrastive sound-effect which cannot be described in terms of the segments, or phonemes, in the sound system of a language, but which extends over stretches of utterances [...]" (ibid.). Schuller and Batliner write that computational paralinguistics build on Crystal's narrow definition of paralinguistics focusing on vocal effects connected to computation, which is how paralinguistics are performed through or in technologies. They write: "we restrict the term to everything that can be found in the speech signal, e.g., in telephone speech or in audio recordings, which cannot be described only in strictly phonetic and/or linguistic terms" (Schuller et al, 2013). The computational paralinguistic is today a consolidated discipline that, in addition to audio recordings, investigates different aspects of computationally automated emotions and personality traits expressed by or embedded in speech and language (Baird et al, 2017; Baird et al, 2018; Meinedo and Trancoso, 2010; Schuller et al, 2013; Schuller et al, 2017; Schmitt and Schuller, 2015).

In this PhD project, I am concerned with the paralinguistic expressions and perceptions of gender, age and accent as well as the affective and imaginary dimensions of paralinguistics. In the field of computational paralinguistics, gender, age and accent are

considered as *long term traits*, more specifically defined as *biological trait primitives* (Schuller et al, 2013, p.7). Biological trait primitives also include phenomena such as height, weight and age which are all understood as long term traits as opposed to *short term states*. Short term states are described as vocal modes which covers the speaking style, the voice quality, and the “emotion-related states or affects” exemplified as e.g. intimacy, interest, uncertainty, frustration, and pain (ibid., p. 8). Within state-of-the-art computational paralinguistics it is the general view that some of the traits and states “can have different intensity, of course, apart from the ones that are binary or can be measured on an interval scale such as age or height” (ibid.). How traits and states are listened to, evaluated and measured is something I will discuss in connection to how the tone of voice and paralinguistic expressions have been considered within sound studies.

Sound studies is a rather new and interdisciplinary field of research. Sound studies scholar Jonathan Sterne has proposed a broad characterization of sound studies as a field that “takes sound as its analytical point of departure or arrival.” He argues that sound studies is an “intellectual reaction to the changes in culture and technology” (Sterne 2012, p. 3) that focuses on sonic phenomena aiming to re-describe “what sound does in the human world, and what humans do in the sonic world” (ibid., p. 2). Sound researchers Trevor Pinch and Karin Bijsterveld write that sound studies is an interdisciplinary field of investigation that is concerned with sound on a broad scale connecting to fields such as media and communication studies, cultural studies and the history and anthropology of the senses (Pinch and Bijsterveld, 2004).

Within the field of sound studies, there has recently been focus on the paralinguistics of the voice as the quantification of vocal tone (Dyson, 2014) as vocal production beyond words (Connor, 2014) and as the affective and performative qualities of voice (Neumark, 2010; Schlichter, 2011; Vallee, 2017). LaBelle writes specifically about the ‘cultures of the paralinguistic’ in his book *Lexicon of the Mouth: Poetics and Politics of Voice and the Oral Imaginary* (2014) in relation to the mouth as the physical site of vocal productions and the paralinguistic auditive aspects of the voice with a focus on the ‘body in the voice’. LaBelle here focusses on the tone of voice as well as on the performative paralinguistic gestures such as whistling, spitting, screaming, mumbling, chewing, or yawning interjecting into conversation. Contrary to computational paralinguistics, LaBelle’s

‘cultures of the paralinguistic’ finds that the communicative meaning of paralinguistics are found in how vocal expressions are performed situated in a social context, a context that includes the oral imaginary. LaBelle sees the oral imaginary as the way the “interior life and the social experience, across geographies and politics [...] lead us into journeys with animals and winged creatures” (2014, p. 183).

In this PhD project, I am not so interested in the ‘cultures of the paralinguistic’ connected to the figure of the mouth, but I am interested in the performance of the contextually situated communicative meaning of the ‘body in the voice’ connected to the oral imaginary, as I will elaborate further upon in the following sections. Just as LaBelle has done, I include words as a material that can be “sculpted, crafted, bent, and produced, as an aesthetical or expressive project” (LaBelle, 2014, p. 173). In my use of technological paralinguistics, I am concerned with how the technological, auditive, performative and imaginary aspects of the situated paralinguistics are connected to experiences of vocal bodies in art and digital media performances.

In the following, I will present how the vocal expression as communicative auditive meaning and as performative paralinguistic body has been conceptualized within the interdisciplinary field of sound studies.

### **Vocal Auditive Meaning**

Sterne explains that the voice has been conceptualized as “the most personalized and most naturalized forms of subjective self-expression” and that the voice is often treated as an expression of subjectivity and consciousness (Sterne, 2012, p. 491). However, the voice as an expression of a present consciousness is something that has been questioned by voice studies scholars ever since the post-structuralist philosopher Jacques Derrida in his book *Of Grammatology* critiqued the metaphysics of presence in relation to speech, writing that the notion of the speaking subject as present was rooted in a phonocentrism connecting voice and being (Derrida, 1967/1997). As Michael Wood, scholar of English and Comparative Literature explains, Derrida argues against the understanding of speech as a direct communication of the presence of the subject (Wood, 1977).

Since Derrida’s critique of speech as act of presence, the voice has become an important field of study within the humanities, and research into the voice has been steadily growing and discussed especially in the last two decades. The philosopher and

feminist Adriana Cavarero, for example, builds on Derrida's critique of the voice as presence and mere being, but critiques his focus on speech by arguing that the voice has an auditive meaning of its own. In the book *For More Than One Voice*, Cavarero writes that the voice is "a communication of one's own uniqueness, that is, at the same time, a relation with another unique existent" (2005, p. 5). This communication is not about accessing anything hidden or essential of a person's subjectivity. The communication should rather be understood as an *interrelationality*, something that opens up a relational space – a plurality – through which the uniqueness is communicated in a polyphony of other singular voices. Cavarero is concerned with "the embodied uniqueness of the speaker as it is manifested in that speaker's voice, addressed to another" (ibid., p. vii). She focuses on the bodily uniqueness of the speaker as it is manifested in the voice of the speaker, speaking to the other and argues that the voice should be understood as a uniqueness defined by virtue of its fundamental interrelationality. The uniqueness of a person exists through an active relationality to other listening and speaking living beings. In this way, it is in the actual action of utterance from one unique existent in relation to another unique existent that 'a communication of one's own uniqueness' is manifested as something that relies on sets of differences in the relation.

In the context of this PhD project Cavarero's theory can point to why the focus on paralinguistic expressions of a speaker are important to discuss, as she argues that what matters in the speech is not the meaningful content or communication. Instead, drawing on the notion of action from the canonical text *The Human Condition* (1958) by the philosopher Hannah Arendt, Cavarero argues that what matters is how one shows who one is by relating one's vocal expressions to others in the moment one speaks as an active interrelational action. This way of understanding the voice does not connect so easily with various forms of identity politics, as it is irreducible to any set of descriptors in that it calls for the recognition of a radical interrelationality.

Cavarero exemplifies the radical interrelationality of voice and why she believes that the auditive meaning of the voice is important by arguing that one can describe Virginia Woolf with a lot of different descriptors; white, lesbian, bohemia, writer, feminist, woman etc., but that these predicates will not capture who Woolf was. Cavarero states that there is a uniqueness of the voice through which Woolf communicates who she is. Something that cannot solely be captured in the predicates alone. Because in Woolf's speech there is

also an active manifestation of who she is in relation to others, expressed as an auditive materiality of a “uniqueness-in-resonance” (Cavarero, 2005, p. 199). By focusing on the way, the unique sound of each and every voice is manifested, Cavarero argues for; “the singularity of the speaker’s voice, the acoustic emission that emits from the mouth to the ear” (ibid., p. viii). Cavarero focuses on the resonant communication of voices in which the speaker’s bodily singularity emerges in relation to the others, no matter what they say. Cavarero’s theory can in this context make clear how the voice and paralinguistic expressions are radical interrelational places, through which one’s uniqueness is communicated in relation to others, not as a passive being but as an active relation and resonance unfolding.

In this PhD project Cavarero’s conceptualization of vocal expression as interrelational communicative auditive meaning has inspired the collective vocal project *Collective Performative Reading* and informed the discussion of paralinguistic multivocality to include a relational perspective.

### **Performing Paralinguistic Bodies**

Neumark writes on vocal performance in “Doing Things with Voices: Performativity and Voice” (2010), arguing that when voices work performatively, they don’t just describe or represent something or someone, but perform and activate, for example feelings of intimacy or intensity. She states; “performativity suggests something that doesn’t just describe or represent but performs or activates” (2010, p. 96). Neumark emphasize that voices can “performatively interpolate” audiences as active performers bringing e.g. artworks into existence (ibid., p. 97). She describes how voices can provoke feelings in the audience as a performative action that enacts through the material qualities of the sounding voice, which goes beyond what the voices semantically say.

Neumark’s description of performative voices connects to LaBelle’s notion of vocal performance. LaBelle builds upon Cavarero’s recuperation of the voice as a sonic materiality “as a process that need not arrive at the semantic” (LaBelle, 2014, p. 62), but argues that Cavarero’s conceptualization of the voice as a uniqueness of active being in relation is also revealed when “we play with words, impersonate others, parade the heterogeneity of our identity” (ibid., p. 62). Here, LaBelle points to the performative aspect of the voice stating; “voice must be emphasized less as an articulation of certainty, and



more as a *performative attempt* at identity” (LaBelle, 2014, p. 141). LaBelle also critiques Cavarero for operating with a voice as a “natural’ property”, which strips away the more “performative dimensions voice comes to weild” (ibid., p. 62). LaBelle writes about the performative voice referring to the philosopher and queer theorist Judith Butler’s seminal theory of performativity as reproduction of norms. He writes that the norms however, can also be subverted through among other things “parody, and clowning – that of things going awry” (ibid., p. 69). This can happen through a performance of clowning of the mouth, not knowing the words or the misbehaving of the mouth. LaBelle exemplifies this referring to the exaggerated performance of gibberish in Charlie Chaplin’s performance in *The Great Dictator* (1940). This kind of subversive performative attempt at identity may incite debate of the existing norms or discourse. The performative attempt of identity through vocal production is always connected to the body according to LaBelle. He writes:

The voice does not move away from my body, but rather it carries it forward – the voice *stretches* me; it drags me along, as a body bound to its politics and poetics, its accents and dialectics, its grammars, as well as its handicaps.  
(ibid., p. 5)

LaBelle here draws on the literary theorist and philosopher Roland Barthes seminal theory of the ‘grain’ of the voice as it is formulated in his canonized essay *The Grain of The Voice* (1977). Here, the singing voice is described in relation to the body as something aside from language. Barthes argues that the “grain of the voice” can be heard as “the materiality of the body” (Barthes, 1977, p. 182). He writes that the ‘grain of the voice’ is a form of bodily communication that circumvents the boundaries of the linguistic sphere and reveals the materiality of the language from within – the body in the voice as it sings. Barthes, in this way, highlights the voice as something attached to the body and writes that the voice is individual, because the voice “has us hear a body” (ibid.). To Barthes the voice is something that is intimately connected to being present, as a confirmation of one’s own physical existence, in that speech can be seen as something that comes from our body as breath and vibrations. LaBelle’s notion of the ‘body in the voice’ connects to Bathes ‘grain of the voice’. However, LaBelle understands the ‘body in the voice’ as a condition where the voice is both subject and object and transgresses the bodily borders, whereby the

classical subjectivity is defined, and lies in the in-between – in a “continual encounter with the other” (LaBelle, 2014, p. x). In this way, LaBelle (building on Cavarero) writes that the voice is one of the most intimate things that can define a person and yet, at the same time, it never completely defines the person. It is always out in the world, as a performative relationality between the voice and the listener.

LaBelle’s ‘body in the voice’ and Barthes’ ‘grain of the voice’ can be related to the literary scholar Steven Connor’s conception of the vocalic body or the ‘voice-body’ in his book *Dumbstruck: A Cultural History of Ventriloquism* (2000). The voice-body also concerns the relationship between the body and the voice. In Barthes’ conceptualization of the relation between voice and body, the body is seen as that which produces the voice. In Connor’s and LaBelle’s theories, there is also an inversion of that relationship. Here, the voices or paralinguistic expressions also produce bodies, physical and imaginary. Connor’s theory of the voice-body suggest that an inversion of body and voice happens constantly.<sup>3</sup> As opera scholar Jelena Novak explains, the voice-body is a mirror mechanism where “the voice is projected by, but also on, the body and that projection, in case of vocal performance, immediately affects the identity and the presence of the body that produced it, by reflecting itself back to it” (Novak, 2015, p. 6). The voice-body can be an imaginary production of a body connected to the voice one hears. As Connor writes this is a body that doubles the body from which the voice emanates. Connor further notes that there can be multiple voice-bodies in the same vocal performance. The voice-body can in this way dissolve dichotomies and cut across separations of the voice as either present or absent, embodied or disembodied, just as LaBelle’s oral imaginary and the paralinguistic expressions of the ‘body in the voice’ can.

Connor writes about the voice-body in relation to the performance of ventriloquism, but I would like to connect it to a discussion of paralinguistic performances developing LaBelle’s mapping of ‘cultures of the paralinguistic’ as well as drawing on Neumark’s conception of vocal performance. Connor’s theory of the voice-body is the foundation for my conceptualization of what I call a performative *paralinguistic-body*. The paralinguistic-

---

<sup>3</sup> Kahn writes that when technologies captures and transmits the voice, it becomes something detached from the body, as disembodiment or de-boned sound (Kahn, 1999). Connor criticizes Kahn’s conception of disembodiment as a violence where the voice is forced away from its body, because to Connor the voice is never without a body implied, it is a voice-body (Connor, 2000). I am here following Connor’s approach to the relationship between voice and body as a voice-body.

body can capture the many different relations between paralinguistic performance and notions of identificatory vocal bodies – including the imaginary of bodies which to me presents a possibility of moving beyond the binary dichotomies of present or absent, embodied or disembodied. This quest of approaching vocal production beyond dualisms is also shared by scholar, sound artist and composer Miriama Young. In her book, *Singing the Body Electric: The Human Voice and Sound Technology* (2015), Young writes on voice and technology, arguing that we should not seek to separate, but blend the synthetic and organic, bodily and mechanical. Vocal technologies are to her both the bodily and digital manipulations and mediations performances (Young, 2015).

In relation to my PhD project, I am also engaged in a quest to listen beyond dualisms, without pretending that they do not exist, and I am therefore invested in listening to the performance of paralinguistic-bodies as vocal material performances of bodies as technological, physical, imagined, human and non-human, sound and data – expanding the boundaries of what is considered a vocal body in art and digital design. The paralinguistic-body is auditive and bodily as Barthes' 'grain' of voice, but it is not operating within the bodily borders of a human subject as Barthes' 'grain' of voice can be said to do. The paralinguistic-body is operating with and beyond the physical borders of the human subject, as paralinguistic expressions that perform the 'body in the voice' (LaBelle, 2014) where the paralinguistic-bodies are continuously in relation and defined in the encounter with the other (as LaBelle is developing Cavarero's theory of vocal auditive difference) including the imaginary bodies (as Connor's 'voice-body') and the non-human in relation to listeners.

In my presentation of the artworks and collective vocal projects in the concluding chapter, I use the paralinguistic-body as a term directing attention to how the material auditive performances create experiences and imaginaries of vocal bodies. This notion of the paralinguistic-body can contribute to the field of computational paralinguistics by focusing on the dynamic interrelationality of paralinguistic expressions and listeners. In the articles included I discuss how vocal expressions are mediated, modified and manipulated through technologies. In the concluding chapter I discuss how these technologies – bodily as well as digital – manifest, transform, alter and complicate notions of identificatory paralinguistic-bodies (as imaginary or real, human or non-human, sound or data) in relation to the listener. I present how technologies can stereotype paralinguistic-bodies. I

also present how technologies of mediation and manipulation can enable paralinguistic expressions that challenge vocal stereotypes and transgresses bodily borders as performing paralinguistic-bodies of entangled partiality, interrelationality, affective multiplicity, and even include nonhuman paralinguistic-bodies in human vocal expression.



1. We only ever speak one language.
2. We never speak only one language.

(Jacques Derrida: *The Monolingualism of the Other;*  
*or, The Prosthesis of Origin*)

**You can recite, shout, sing, yell,  
scream, murmur, mumble, or  
whisper the words you read using  
headset, mouth, microphone,  
larynx, vocoder, vocaloids,  
synthesized speech**





# Listening to and through Voices

Listening – and listening to voices – is something that has been extensively theorized within sound studies. Sterne writes that in the interdisciplinary field of sound studies theorists, artists, scientist and scholars are studying sound and listening through both theoretical and practice-based research in ways that “thinks across disciplines and traditions, some that have long considered sound, and some that have not done so until recently. Sound studies is an academic discipline, but it can also move beyond university” (2012, p. 2). Weidman writes that it is through the way in which the voices resonate in listeners and publics that the meaning becomes audible:

Voices are constructed not only by those who produce them but also by those who interpret, circulate, and reanimate them: by the communities of listeners, publics, and public spaces in which they can resonate and by the technologies of reproduction, amplification, and broadcasting that make them audible (2014, p. 45).

She writes that the voice is both a matter of expression and of being heard. This aligns with Lawy, who states that the voice is not only about “the act of speaking” but also about listening and “the situation and context in which it is made” (2017, p. 194).

In my PhD project, I have used many different approaches and methodologies researching the performances of paralinguistic-bodies. But, like many others in the field, my predominant approach has been to listen. I do not only listen to voices as a method, I also co-create new vocal sounds to listen through. I approach the paralinguistic-bodies in art and digital media through methods of listening to and through voices.

In the following, I shall first explain my approach of listening to and through voices as something that is interpretive, embodied, situated and relational, and then I will unfold the method of collective practice-led research as a way to listen to and through vocal bodies as performative paralinguistic expressions.

## **Interpretive, Embodied, Situated and Relational Listening**

Listening means to take notice of or to pay attention to sounds, underlining the deliberate action of comprehending what is heard. The aim of my PhD project is to investigate the listening experiences of voice and vocal bodies in artistic work and digital media production. In sound studies, there are many ways to pay attention to sounds while listening. I have approached listening as a mode of interpretation, as an embodied and situated experience, and as a relational and performative act that includes the listener's interpretation in the understanding of what is being heard. In the following, I will run through some of the seminal texts by scholars in sound studies unfolding these different approaches to listening that I account for more or less explicitly in the articles and collective vocal projects part of my PhD project.

Listening as an act of interpretation is predominant in the composer, music educator, environmentalist and founder of soundscape studies Raymond Murray Schaefer's numerous listening terminologies. Here the interpretation of soundscapes is supported through terms that help decipher sounds into different categories, e.g.; keynote sounds, signal sounds and sound marks (1977). Barthes also operates with listening as interpretation in his essay "Listening" (1985), in which he writes that listening is where the ability to interpret and decipher sonic signals into other perceptual territories begins. The composer and writer Michel Chion also writes on the how to interpret sound through different listening modes focusing specifically on audio-visual relationships. He presents three listening modes (causal, semantic and reduced listening) as different ways of directing auditory attention to sounds in terms of how they open for interpretation (Chion, 1990).

In the book *Listening and Voice: Phenomenologies of Sound*, the philosopher Don Ihde develops a theory of embodied listening in which he takes the situated bodily engagement as point of reference for the experience and interpretation of voices and "the voiced character of the sounds of the world" (1976/2007, p. 147). Ihde addresses here the polyphony of sense perception where all senses are part, but wherein listening to voices holds a special position (ibid., p. 116). The experience of the world's voices is through actions created by human or non-human agents through bodily movements – as embodied perception or listening. Embodied listening is when the whole body and all the senses are engaged – a perception that can be mediated through (technological) artifacts (ibid.). Ihde

understands listening as an embodied and therefore situated, material activity. Schulze adds that listening is all this but also a relational, situated, material and performative activity in which the “various layers of the body of the listener” as well as the surrounding contexts “the walls and the floor” and the technological equipment and historical situatedness matter to how listening is done (2018, p. 34). Schulze’s conception of the listening activity as situated in materialized, and technological relations connects to the feminist theorist and biologist, Donna Jeanne Haraway’s conception of ‘situated knowledge’. Haraway develops her concept of situated knowledge in her essay “Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective” from the book, *Simians, Cyborgs and Women* (1988). Here she writes that knowledge comes from a body, a body that is partial, subjective, and situated. Knowledge that comes from a specific body, also comes from a specific place and time, and is therefore not objective or universal. In this way Haraway’s situated knowledge challenge the myths of traditional disembodied objectivity. Objectivity, she writes, is in fact always already partial, embodied, and specific. Haraway argues that one way of drawing attention to the politics of knowledge production as something that is not disembodied or universal is to make clear that vision is partial, positioned, embodied and located. Haraway writes that the bodily situated partiality “is the condition of being heard to make rational knowledge claims. [...] the view from a body, always complex, contradictory, structuring and structured body, versus the view from above, from nowhere, from simplicity” (ibid., p. 589).

Haraway’s situated and partial subject producing situated knowledge resonate in LaBelle’s notion of situated and located listening. He writes that “the physicality characteristic of sound suggests a deeply impressionistic, locational ‘knowledge structure’ – that is, the ways in which listening affords processes of exchange, of being in the world, and from which we extend ourselves” (LaBelle, 2012, n.p.). To LaBelle, situated listening is something always nested within a relation, something which creates an emerging sociality and radical sharing:

to *give* one’s ear is to invest in the making of a future public; it is to give the body over, for a distribution of agency. [...] A listening that gains in momentum precisely by integrating the productions inherent to auditory experience, expressed in the migration of voices, the shifting of the body, the

animation of knowledge, as well as the deepening of attention; in short, the production of radical sharing. (LaBelle, 2014, p. x)

Here, listening is not only connected to interpretation but also to a distribution of agency as a radical embodied sharing. LaBelle's conception of listening as a process of exchange and as a radical sharing connects to the philosopher Jean-Luc Nancy's notion of listening as a way of approaching the self-in-relation, writing: "When one is listening, one is on the lookout for a subject, something (itself) that identifies *itself* by resonating from self to self, in itself and for itself, one in the echo of the other" (2002/2007, p. 9). Nancy, like Cavarero, builds on Derrida's conceptualization of difference and can, in my opinion, be seen as investigating radical relationality from that of listening instead of the voice. To both Nancy and Cavarero, listening (to the voice) is about relationality and resonance. Cavarero, however, argues that Nancy's philosophy of listening "does not pay attention to the materiality of the acoustic sphere" (2005, p. 90), and that Nancy has a passive concept of listening and being-in-relation, whereas Cavarero has an active interrelational concept of listening and sonorous speaking.

In the context of this PhD project the above insights into listening as interpretive, situated in the body as well as in relation to others also means that different listeners – when listening to the same paralinguistic expression – experience different things. This experience is depending on "their spatial orientation as well as their physical, cultural, social and historical predispositions" as Rasmus Holmboe explains (2019, p. 4). My listening experiences of the paralinguistic expressions discussed in the articles and collective projects included in the dissertation will never be exactly the same as somebody else's experience, but are informed by my situatedness.

Though I attempt to show more or less explicitly how I as a listener am informed by my own situatedness, desires, motivations, relations and interpretative interest, the complexity of how all that goes into listening may never be fully recorded in my analysis and discussion of the technologically mediated and modified paralinguistic-bodies in artworks and digital media performances. However, what I hope to make clear here is that I am a listening subject who strives to listen from a specific place with particular interest and investment.

The discussion of the vocal bodily performances as they are presented in the articles and vocal collective projects, are informed by my investment and particular interest in listening to and through voices as performative paralinguistic expressions, which has been developed over the years. My interest in vocal performances really took shape as my *project*<sup>4</sup> when I joined *Tonespace* in 2005, an educational program in electronic music and sound installation, at the Danish National Academy of Music. Shortly after, I began studying sound- and performance art at University of Copenhagen. Here I investigated sound- and performance works through a practice-led approach. First, I explored the iconic work by Alvin Lucier *I am Sitting in a Room* (1969) through practice-led research (Smith and Dean 2009), investigating my own experience of listening to and through voices, remaking it in what became *I am Sitting in a Different Room* (2010) and *Jeg Står i et Rum* (2010). This led me through a process of recording and re-recording (and re-re-re-recording). Here, instead of re-recording and playing back the sound of my voice in a resonant chamber, like Lucier did over and over again in different locations, I recorded and played back my voice in a controlled reverberant chamber and in an anechoic chamber (a room with almost no resonance) at the Danish Technical University (DTU) in order to hear how the different spaces and the technologies in it would affect the texture of tone of my voice. My experiment in the anechoic chamber did not bring out any resonant frequencies of the tone of my voice or my bodily sounds, but instead brought out what the artist and researcher, Douglas Repetto, calls “the technological resonances of the equipment” (2011, n.p.). What is heard on the track is the noises from the microphone and computer in the room made audible in relation to the tone and rhythm of my vocal expressions. My listening of the voice was entangled with the technological resonances. I wrote about my practice-led research approach as a way to investigate “the process unfolding in Lucier's piece,” and as a way “to explore different recording technologies and to manipulate the voice with the use of different spaces” (Hasse, 2012, n.p.).

My studies have consistently relied on practice-led research, researching vocal material performativity through series of practical experiments. Most of my practice-led research have been concerned with the exploration of technological manipulations of sonic qualities and meanings of voices in relations to listeners, technologies, bodies, contexts

---

<sup>4</sup> Mary Kelly writes, in the introduction to *Dialogue - On the Politics of Voice* (2012), on project-based works as something that is developed throughout very long periods of time through many smaller projects.

and situations. My interest has led me to study sound production and sound design at Royal Holloway, University of London (2008-2009) and later to study at Columbia Computer Music Center, Columbia University where I experimented with vocal performance technologies (2010 and 2015). My interest in the voice as a bodily, sonic materiality, as well as a technologically mediated, manipulated and performative expression, has developed throughout the years. As a practitioner and researcher of vocal production, I have been increasingly interested in listening to and through vocal expressions studying the performative and affective dimensions of voice and multivocal paralinguistic practices. Listening is a way for me to include my situatedness as a listener (with the experiences, interests, relations and investments described above) in the interpretation and performance of how vocal expressions come to matter as paralinguistic-bodies.

### **Performative Listening**

In my listening to and through paralinguistic expressions, I have also been inspired by how art historian Camilla Jalving approaches artworks as performances in her book, *Værk som Handling (Work as Action)* (2011). Jalving develops her methodological approach in connection to the language philosopher John Langshaw Austin's speech act theory published in the seminal book *How to Do Things with Words* from 1962. One of the important elements in the speech act theory is Austin's development of the *performative sentence*, defined as a sentence that does not only *describe* something, but as a sentence that *acts*. "To name a ship *is* to say (in the appropriate circumstances) the words 'I name, &c'. When I say, before the registrar or alter, &c., 'I do', I am not reporting on a marriage: I am indulging in it" (1962, p. 6). In the quote, Austin argues that the words do not just describe a situation, but creates situations through performative sentences – sentences that acts. Or as performance theorist James Loxley explains, "the utterance is not setting out to *describe* a situation, an event or an action: it *is* an event or action" (2007, p. 8).

Jalving uses Austin's theory to argue that art can be understood as action and not just descriptions (or objects) when placed in the right social contexts and circumstances. She writes about how a contemporary artwork does something in that it stages, actualizes, dramatizes and performs itself in relation to its audience. Jalving argues that viewers of an artwork also witness the artworks action, and that it is therefore important to look into not

just *what* the artwork is, but also *how* it is. The performative sentences such as “I do” or “I name this ship” become actions if they are uttered in the right context under the right circumstances. When “I do” is said in the city town hall or in front of the alter in the church, where a mayor or a priest conduct a ceremony and can act as a witness, the utterance becomes a performative sentence, since the utterance of the words “I do” results in a couple being married. Speech act theory does not only focus on the sentence as an action, but also on the social conditions which can confirm the sentence as an action.

In my analysis of works in art and digital media, I focus on the vocal paralinguistic expressions as acts and as something that I as a listener takes part in through my continuous interpretive, embodied, situated and relational listening – through my continuous ‘performance’ of the act of listening in relation to the vocal performance in the work. To paraphrase Jalving’s description of the performance of visual art; the answer to what the work *is*, is that the work is what it *does*. The work is its action. Therefore, as a listener, I should not only orient myself towards *what* the artwork is about, but also *how* it is performed – in this case how the vocal expressions perform paralinguistic-bodies, which includes me as a listener and my interpretation of the vocal performance in the work. The *doing* of the vocal performance in an artwork is, in this way, closely connected to my act of listening as a listener.

The difference between the utterance “I do” and the voices as utterances in the contemporary artworks that I am concerned with in the articles included, is that the performative sentence is created in relation to its listener not solely through *what* is said semantically, but also *how* it is expressed paralinguistically. It is through the materiality of the voices, their intensities and affects, that the listeners are interpolated as witnesses and active participants in the creation of the artworks’ performative sentences (not as descriptions of a situation, but as actions that include the listener), as Neumark has stated. I will elaborate on the ways in which the vocal intensities and affects in the artworks produce paralinguistic-bodies to me as a listener in the conclusion. Here, I also argue that by taking the performance of listening into account when discussing the auditive expressions of technologically generated and mediated voices in artworks, but also in relation to other sites for vocal expression, is something that can develop e.g. the field of computational paralinguistics (to include the performance of listening into the analysis of paralinguistic expression).

## Practice-based and Practice-led Research

One of the ways in which I have performed listening is through the vocal paralinguistic expressions created in the *Collective Performative Reading* and the *[multi'vocal]* project. In the two practice-led research projects I focus on the collective vocal paralinguistic expressions as something that I as a listener can take part in through my continuous embodied, situated and relational listening. My situated exploratory approach is very much inspired by LaBelle's methodological approach of listening from the position of *not knowing* – “to listen is to adopt a position of *not knowing* [...] listening is an unsettling of boundaries – what draws me forward away from what I know” (2014, p. x). In the practice-led research projects I have used listening as a starting point for *not knowing* – as a method to investigate the possible meanings of the performative auditive (and technologically mediated, manipulated and synthesized) paralinguistic-body.

Within the field of sound studies specifically, humanities and contemporary arts at large, practice-led research as a subgenre of practice-based research has been growing as a research methodology and is becoming more and more consolidated. Furthermore, it has been met with growing acceptance within academies and university environments. Scholars of practice-based research approaches Linda Candy and Ernest Edmonds writes that a “basic principle of practice-based research is that not only is practice embedded in the research process but research questions arise from the process of practice” (2018, p. 63). They write that in creative arts “including new media arts, the emphasis is on creative process and the works that are generated: [...] practice and research together operate in such a way as to generate new knowledge that can be shared and scrutinized” (ibid.). Candy writes in another text on practice-based research, that when the practice is an integral part of the research process, the outcomes or process of the practice needs to be documented and reflected upon in order to support its position in the overall research project (2006, p. 2). In the context of this PhD project the texts on *[multi'vocal]* and the *Collective Performative Reading* document some of the concerns and considerations connected to the projects, which are also reflected upon in relation to the common research investigation on paralinguistic expressions and vocal bodies in the concluding chapter.



In articulating how practice-based research can be unfolded (as a sub category of practice-based research), scholars Hazel Smith and Roger Dean write, in *Practice-led Research, Research-led Practice in Creative Arts* (2009), about what makes practice-led research distinct. In this PhD project, I have a practice-led approach, which according to Smith and Dean is a hybrid approach between creative practice and academic research based on a belief that knowledge is “often unstable, ambiguous and multidimensional, can be emotionally or affectively charged, and cannot necessarily be conveyed with the precision of a mathematical proof” (2009, p. 3). However, Smith and Dean define practice-led research as a form of research in which the practical work generates research insights that can be documented and theorized. The dialogue between theoretical investigation and analysis generate research insights that can inspire the development of the practical dimension of the project and vice versa. Smith and Dean argue that this kind of research can be seen as performative. In this way, the collective practice-led research projects can be understood as performative acts, in that the practical projects presented, are viewed as important elements in the investigative process and not just a description of it. Here the vocal projects should be understood as manifestations of a knowledge that is practiced and performed through their practice-led research *modus operandi*. The remediation of the *Collective Performative Reading* included in the PhD dissertation as quotes and instructions can be seen as an attempt to perform elements of the research and not just describe it.

The artist-philosophers, Erin Manning and Brian Massumi, discuss the entwinement of academic research and creative practice with the concept of “research-creation”, a term meant to “encourage hybrid forms of activity” in an interdisciplinary field between art practice and academia (2014, p. 84). In recent years, new approaches to do research-creation within the field of sound studies have emerged, which have added many different entry points for researching sound. Artist-researchers, like Young, Labelle, Cathy Lane or Lawrence Abu Hamdan, work with exploring vocal and sonic phenomena through their artistic works as well as through theoretical investigations. Young considers electronic voices through new music compositions, Labelle examines voice and sound through events and installations, Lane explores voices and the spoken word through compositional approaches and Hamdan investigates how forensic listening to voices have created new vocal subjects through artistic audio documentaries and installations. Common to these

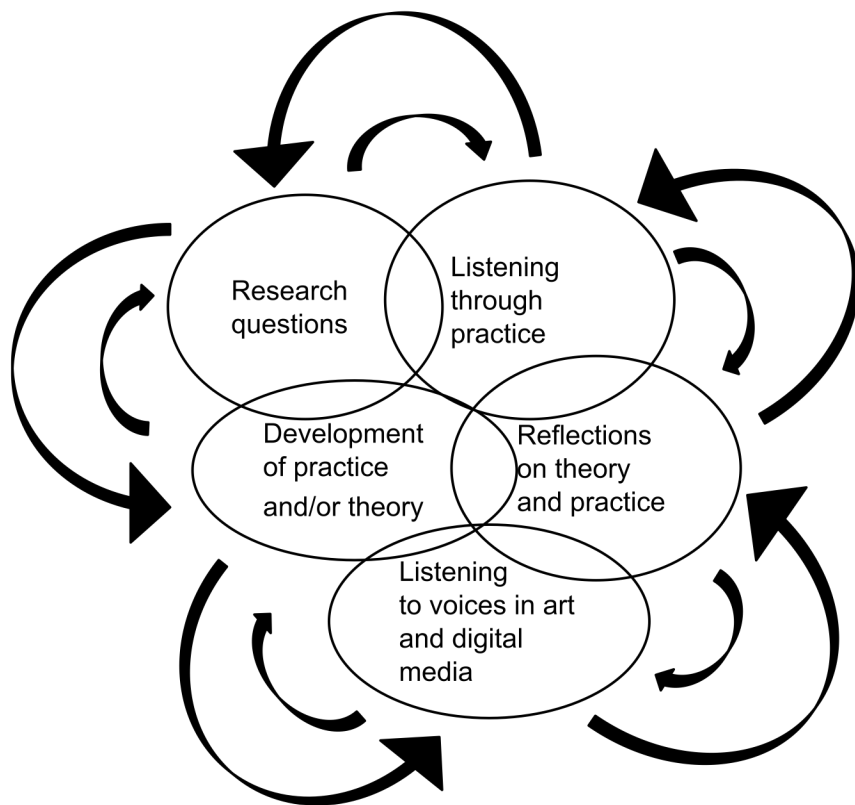
artistic engagements with the voice is that they introduce alternative ways of approaching the voice opening for new philosophical questions and discussions. These approaches are aligned with the contemporary field of artistic research. Art- and design theorist Kathrin Busch has characterized artistic research as a constant negotiation between academia and artistic practice where artists “integrate research methods and scientific knowledge into their artistic process to such a degree that it even seems to be developing into an independent form of knowledge on its own” (2009, p.1).

In many ways, my research practice links to the artistic research presented by Hamdan, LaBelle, Lane and Young, in that I too explore vocal phenomena through inquiries in my creative practice and my theoretical work. However, I often do so in collaboration with others. My explorative practice-led investigations have been created with other researchers, artists, designers and programmers. In the project *Move/Bevæg Dig* (2012), I collaborated with composer and instrument builder Jeff Snyder to explore how the relations between bodily movements in space and vocal sonic expressions are enabled through ultrasonic distance sensor technology and how the experience of generating vocal sounds (and possibly words) by moving the body in a specific room reconfigure the relationscape between voice, body and space (Jørgensen and Snyder, 2012). I made the installation *Electromagnetic Landscape* (2015) together with visual artist Ayaka Okutsu and artist/programmer Daniel Cermak-Sassenrath that explored how radio (voices, music, noise etc.) could be created through the connection between the bodily actions of multiple participants using electromagnetic technologies (Cermak-Sassenrath, Okutsu and Hasse Jørgensen, 2015). Together with Cermak-Sassenrath, I have recently created the project *Re-volution Sampler* (2018). Here we explore the possibilities and challenges of creating a participatory sonic archive enabling participants to explore and engage with historical revolutionary songs through their own singing – creating a layered vocal soundscape of people singing the same song throughout time (Cermak-Sassenrath and Hasse Jørgensen, 2018). In this way, I have had a long-time investigation of the relation between listeners, technologies, bodies and performative vocal auditive expressions through different collaborative practice-led exploratory vocal projects.

Common to the different collaborative practice-led projects is that my interest does not lie in discussing the authority of the work. Instead, I look into the generative potentials of practical work as a way of investigating a specific aspect of voices in practice and theory

using “the multidimensional, reciprocal and iterative relationship between research and practice” (Smith and Dean, 2009, p.10).

My work with practice-led research is based in an understanding of practice as something that can generate research insights, which then can be performed, documented, theorized, and reflected upon. In connection to Smith and Dean’s framework my practice-led research is unfolded through an *iterative cyclic web*, meaning that the work is process-driven, and that it does not have any particular starting point or any conceived end-point as such, and that the plan for this practice-led research is open for transformation.



The iterative cyclic web of the practice-led research in my PhD project.

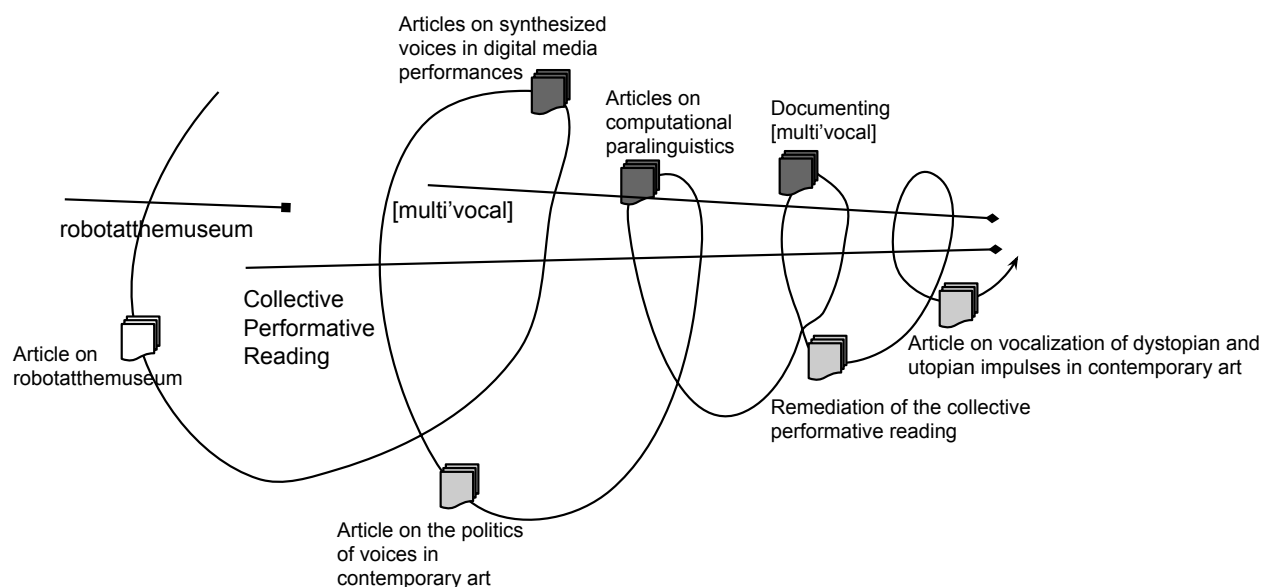
The iterative cyclic web serves as an overall frame for the approaches that I have used in the development of the whole PhD project. In my version, the iterative cyclic web combines three different elements:

1. **the cycle** of alternations between practice and research

2. **the web** where there exist numerous points of entry and cross transits within the practice-research cycle

3. **the iterations** through which the different cycles of the development of the creative practice inform the theoretical research that again inform the creative practice pointing to new aspects of the practice to be unfolded.

The cyclic structure is web-like and relational and it is possible to jump from one point in the cycle to any other point. For example, can a publication with conceptual research lead to the revision of ideas in the creative practice that can then inform new experiments or theoretical work within the project. The iterative cyclic web model allows for a depiction of a process-driven way of working, which can lead to ideas that are unforeseen at the beginning of a practice-led project. The documentation and self-description of the creative work and the growing self-critical awareness plays a crucial part in the evaluation of the practice-led project. There is a processual evaluation inherent in the iterative cycles through which peer-reviewed articles, exhibitions and talks at events and galleries are developed in the practice-led research. In my PhD project the practice-led research unfolds as a cyclic web-like process in the attempt to listen to and through the performative auditive paralinguistics of voices from multiple perspectives.



The cyclic spiraling process and outcomes of my PhD project.

In the model above I have mapped out some of the outcomes related to the process developing this PhD project. The cycles are formed in a spiral in order to demonstrate how I become more and more specific and focused with each cycle. The model shows how I slowly specified the research question, through explorations in practice and theory, that then shaped my focus in my theoretical perspective and on the collective vocal projects. The practice-led research process in my PhD project can, in this way, be seen as a cyclic web-like spiraling project oscillating between and merging practice and theory.

### **Collective Practice-led Research**

Smith and Dean write that often practice-led research is carried out collaboratively, where a group of collaborators or several different people may be involved at different stages in the process of developing the creative work and research. In my case, I am not only collaborating with different partners developing the practice-led research, I am working in two art-based collectives: *For More Than One Voice* and the *[multi'vocal]* collective. I have developed my practice-led research as a member of these collectives in connection to the vocal projects collectively created. In the following, I will present the two collective vocal projects and the ways in which the collective projects informed my way of thinking and working with vocal expressions. Then I will broaden the scope of what collective practices can mean in a broader political context.

Common to the two ongoing collectives, *Fore More Than One Voice* and *[multi'vocal]*, is that they are both participatory vocal projects. In relation to my research process, the collective vocal projects have been developed through an ongoing interplay between my academic writing (my analysis of vocal performances in art and digital media) and the collective practical explorations. Here I have been exploring by listening to and through the vocal bodily expressions performed in these projects. This has drawn my attention to the different ways in which voices can perform paralinguistic multivocality as heterogeneous forms of relationality that transgresses the individual as resonance created in relation to others and listeners.

In the collective *For More Than One Voice* of which I am part of together with artist Jane Jin Kaisen, we organized collective readings, in which we investigated the performative,

sonic and affective soundscape of multiple voices by reading aloud together. The *Collective Performative Reading* was, among others, inspired by the concept of vocalic interrelationality developed by Cavarero and the concept of speech acts as they have been theorized within linguistic theory. The collective readings also resonate with Butler's writing about the importance of bodies in "orchestrated collectivity" where assembling or "acting in concert" can be an embodied form of "exercising a plural and performative right to appear, one that asserts and instates the body in the midst of the political field" manifesting "plural forms of agency" (Butler, 2015, pp. 4-11). The collective readings aimed to create a space where affective paralinguistic expressions are political in that they exercise 'a plural and performative right to appear' as a multivocality of interrelated bodies. We also investigate listening to vocal expressions as a situated, embodied interrelationality. Listening to paralinguistics were explored as a resonance from self to self 'in the echo of the other' and as a situated process of exchange from which the self is extended beyond its bodily borders to include the echo of the other as an 'emergent sociality' and a 'radical sharing'.



Organizing the cards with quotes and testing Nao performing quotes at NLH Space, 2016.

In the first *Collective Performative Reading* (out of three so far) at the gallery *NLH space* (shown in the picture above), we experimented with the social robot Nao. We programmed Nao to produce boyish sounding synthetic voice to sonify quotes alongside the vocal performances of quotes from the human participants. Nao's performance of the quotes felt like it interrupted the shared vocal performance-flow. It felt like the robot was not listening to collective auditive space as it was performed and did not integrate with the vocal polyphonic collective space of the other participants listening and performing paralinguistic bodies. It was a disconcerting moment (Verran, 1999). Or put another way, it was a distressing moment with generative potential. It pointed out to me the embodied and relational aspect of the *Collective Performative Reading*. The boyish paralinguistic expressions of Nao felt too boring with a voice that seemed "disaffected", as Dyson calls it (2014), despite having programmed the voice to speak with different pitch and tempo.

In relation to the project *robotatthemuseum*, I had already started to see if I could make the vocal performance of Nao more interesting by inserting human voices reading

aloud manuscripts to be played back when the Nao robot performed. I here started experimenting with making Nao perform multiple recorded human voices in the same performance.<sup>5</sup>



Participants recording different voices to be performed by the Nao robot, Columbia Computer Music Center, 2015.

The project of designing a voice performing with multiple recorded voices started the collective *[multi'vocal]*, which I created with sound artists and PhD researcher Alice Emily Baird, creative programmer Frederik Tollund Juutilainen, interaction designer Nina Højholdt and programmer Mads Pelt. *[multi'vocal]* is a participatory critical design project which consists of two interrelated elements: a recording booth in which participants-as-speakers (and imaginary listeners) are invited to pronounce a sequence of words, and a

<sup>5</sup> See the vocal experiments on the website connected to the project *robotatthemuseum* (2015 - 2016) and the article "Technological Fantasies of Nao – Remarks about Alterity Relations" (Jørgensen and Tafdrup, 2017) included in the appendix. I made the project together with Oliver Alexander Tafdrup at the Medical Museion at the beginning of my PhD project.



website where participants-as-listeners (and potential speakers) are invited to listen to the words created from all the participating speakers. The *[multi'vocal]* investigates among other things how to create paralinguistic multivocal expressions of a synthetic voice as a radical alternative to existing commercial synthetic voices, which perform vocal stereotypes. Currently, most listeners to whom we have presented the *[multi'vocal]* voice, perceive of the new synthetic voice as fragmented and hard to categorize as male or female or as having one specific accent or age (Juutilainen, 2019).

*[multi'vocal]* was inspired visually by Walter Pichler, *TV Helmet* (Portable Living Room) from 1967 and can be understood in relation to the feminist Human Computer Interaction (HCI) design fiction project, the digital personal toilet assistant *U* – in which designers experimented with different voices having diverse ages, genders and accents (Søndergaard and Koefoed, 2018) – as well as to *Q* (2019-) and *Common Voice* by Mozilla (2017-).

As an open source, participatory and non-binary gender alternative to the existing synthesized voices, *[multi'vocal]* prompts reflection and discussion on the current designs of synthesized voices in the everyday life. With the *[multi'vocal]* synthetic voice participants are invited to listen to and through their paralinguistic expressions. Listening for the resonances in the partialities of connection across differences in the multivocality heard. *[multi'vocal]* can act as a catalyst for debate and discussion about what kinds of futures participants want. It is the debate and critical reflection, which is the main aim of *[multi'vocal]*. The recording booth have been installed at festivals (e.g. Roskilde Festival 2017, Techfestival 2017), public institutions (e.g. IT University of Copenhagen 2017, IDA 2017-2018) and gallery spaces (Catch 2018 - 2020). In the text, we explain the process of collecting voice samples from a diverse group of participants and discuss possible futures developments of the project.



Installation of the *[multi'vocal]* voice recording booth at Roskilde Festival, 2017.

The *[multi'vocal]* project has been inspired by *robotatthemuseum* working with synthetic voices and the *Collective Performative Reading* exploring the role of the social robot and the synthetic vocal performance in the polyphonic vocal space collective created by reading and listening to quotes written on cards (which is remediated in this dissertation).

### **The Political Context of Collective Work**

Collective means ‘done by people acting as a group’ and relates to words such as ‘common,’ ‘collaborative,’ ‘communal’ and ‘shared’. Although collective process-based work is not new within the art and design, the collective as basis for practice-led research crossing over between creative and academic work, is rather new. As late as in 2005 performance theorist and maker Bojana Cvejić wrote that collectivism is seen as an outmoded term connecting to theatre and performances in the 1960’s and that it does not appear as a viable model of experimentation and critique to many within performance arts and theory (Cvejić, 2005). This has changed. In relation to a broader political context, the two collective vocal projects, *Collective Performative Reading* and *[multi'vocal]*, can be seen as tied to a larger movement of artistic and cultural collectives (*Collective/Kollektiv*, forthcoming 2020) within the last five to ten years. Such collectives have been formed as counter-reactions to the neoliberal entrepreneurial promotion of a particular kind of individualism wherein “individuals ought to care only for themselves, and not for others” (Butler, 2015, p. 12). This mind set is related to working conditions as the “post-Fordist

forms of flexible labor that rely on the substitutability and dispensability of working peoples” (ibid., p. 11), creating precarious living conditions for most people, including me and my fellow collectivists.

The philosophers and writers Pierre Dardot and Christian Laval argue that neoliberalism changes the way we relate to ourselves and to others, because it “aligns social relations with the model of the market; [...] it even transforms the individual, now called on to conceive and conduct him – or herself as an enterprise”(Dardot and Laval, 2013, p. iii). As cultural researcher Sarah Charalambides writes, human beings are rendered as human capital or private enterprises where they need to self-invest in order to enhance their future value: “This ‘Me Inc.’ model seems to destroy social relations because it replaces them with a purely competitive logic” (Charalambides, forthcoming 2020). In this light, the two collectives can be argued to function as resistance and recuperation of social relations, promoting the collective work and not the individual work. The collectives should not be understood as a site of unity but rather as a site for relation of shared difference. As Charalambides writes, this entail a new understanding of a ‘collective’ as not “a harmonic union of people but a constant investigation of what is shared and how the commonalities manifest, as well as how to include the differences between people without losing the idea of the collective” (ibid.). Charalambides argues, drawing on Haraway, that instead of presenting a unified subject as a prerequisite for the collective, we should see the collective as constituted by fragmented and partial subjects that are changing in relation to others. According to Charalambides, this enables an overcoming of the dichotomous distinction between the ‘I’ and the ‘we’ in the collective “building solidarity across difference(s)” (ibid). Charalambides proposes that the collective can be seen as where resonances happen, in “a co-essence in which the subject is never purely self-referential. It is not ‘me’, neither ‘the other’, but always the result of a resonance” (ibid). Charalambides writing here resonates with what curator and art theorist, Irit Rogoff writes in the article “WE: Collectivities, Mutualities, Participations” (2002). Here, Rogoff looks for a new notion of “we” – as something that connects to shared production of meanings in relation to temporarily formed participations in and around an artistic work. Participation is meant here as a performative function and the “we” can, as Rogoff writes be understood as “emergent possibilities for the exchange of shared perspectives or insights or subjectivities – we allow for some form of emergent collectivity” (2002, p. 127).

She labels this kind of collectivity as a “performative collectivity”, a term which creates an alternative view on “collectivity” and “community” as it is normatively defined upon geographic or ethnic kinship. Rogoff instead proposes to view collectivity as “produced in the very act of being together in the same space and compelled by similar edicts,” something which, as Rogoff argues, “cannot be recognised in the normative modes of shared beliefs, interests or kinships” (ibid., p. 127).

### **Practice-led Research and Collective Feminist Work**

Practice-led research that relies on collective work with others also means entrusting fellow partners with the capacity to alter, elaborate, and change the perspectives of my research by showing new directions or bringing other positions into the shared space of our collective work. My original research question has been expanded and transformed to resonate with other ideas and aims, from which it gains its continuous relevance in the collective. In Rogoff’s words, my research has also been shaped by the ‘exchange of shared perspectives or insights’ created through the collective work. This highlights that the research developed throughout this PhD project has not solely been created or motivated by me, but owes a great deal to my fellow collectivists. This does not entail a unison collective perspective, but rather, as Charalambides states, as a site for relation of shared difference.

Feminist theorists also bring forth critiques of individual autonomy in order to reframe social relationality. Both *[multi’vocal]* and *For More Than One Voice* are feminist collectives inspired by thinkers such as Butler, Haraway, and Cavarero. Common to these thinkers is their emphasis on the inherent partiality, multiplicity, heterogeneity or entanglement of collectivity, while calling into question the collapse of various experiences and situations into one stable and undivided subject position. They bring a re-theorization of individuality that does not refer to an “auto-logical and self-contained individuality”, but rather to a “becoming-with-one-another” manifested in affects such as “indignation, despair, desire, outrage or hope” that are shared by others (Butler and Athanasiou, 2013, p. 71). These shared affective experiences and articulations can interrelate partial subjects and produce an alternative to the neoliberal ‘Me Inc.’ model.

Following Haraway, the partial is locatable and allows for “the possibility of webs of connections called solidarity in politics and shared conversation in epistemology” (1988, p.

584). With regards to collective organization and building solidarity across difference(s) in neoliberal capitalism, the notion of partiality allows for thinking of a collectivity in which identities perform beyond binaries (that signify relationships of 'either or'). Once again, Cavarero's conception of voice as something always in relation to other voices is relevant. Because here, vocal expression is not about accessing anything hidden or essential of a person's identity, but should rather be understood as something that open up a relational space – a plurality – through which the uniqueness is communicated by a polyphony of other singular voices (2005). From this perspective, listening to and through voices becomes vital. Listening for the partialities of connection across differences. This is what we explore in *For More Than One Voice* and *[multivocal]*.



Ne dire mot: tout juste *m* ou *mu*, *muttio*,  
*mugio*, mugir, *mûnjami*, *mojami*.

(Jean-Luc Nancy, *À l'écoute*)

**You can gasp, cough, burb, grunt,  
(whistle) sigh, pause, or yawn as you  
like**





## The Articles

Common to the three articles included in the combined PhD dissertation, *ROBOT* and *THE END* (2018), “Stemmernes Politik i Samtidskunsten” (The politics of Voices in Contemporary Art) (2019),” and “Vocalizing Dystopian and Utopian Impulses” (in press, 2020), is that they all concern the voice as an artistic medium in contemporary art and digital media performances.

In the first article, “*ROBOT* and *THE END*: A Comparative Critical Reading of the Staging of Synthesized Voices in Digital Media Performances”, I discuss the digital media performances *THE END* (2012) composed by Keiichiro Shibuya and *ROBOT* (2013) choreographed by Blanca Li. In the second article “Stemmernes Politik i Samtidskunsten” (The Politics of Voices in Contemporary Art), I discuss *Emigrant* (2013) by Niels Pugholm, *Miss no. 1,2,3* (2013) by Nanna Lysholt Hansen, and *The Woman, The Orphan and The Tiger* (2010) by Jane Jin Kaisen and Guston Sodin-Kung. In the third article, “Vocalizing Dystopian and Utopian Impulses”, I discuss Wagechi Mutu’s *The End of Eating Everything* (2013), *Transgressions* (2001) by Nalini Malani and lastly Nanna Lysholt Hansen’s *Dear Daughter/Anatomy of the Chthulucene #6* (2018).

### **ROBOT & THE END: A Comparative Critical Reading of the Staging of Synthesized Voices in Digital Media Performances (2018)**

“Bodies of Difference/Kroppe i forskellighed,” in *Peripeti – Tidsskrift for dramaturgiske studier*. Eds. Karen Vedel, Jeuno JE Kim, Storm Møller Madsen, Nina Cramer. Årg. 15 Nr. 29/30, 2018.

The comparative critical reading and analysis of the two digital media performances presented in the article, *THE END* (2012), composed by Keiichiro Shibuya and *ROBOT* (2013) choreographed by Blanca Li, concerns my situated listener experience of the staging of synthesized voices in two digital media performances. In the article, I focus on how the use of synthesized voices in the digital media performances operates with the voice as a marker of identity in different ways. As I argue, the piece stages the synthetic voices as

markers of identities manifesting the voices heard either as fetish object or caricature. In many ways, this article sets the scene for my inquiry in the PhD project, in that the article describes how synthesized voices are in my experience designed within a gender binary framework.

### **Stemmernes Politik i Samtidskunsten (The Politics of Voices in Contemporary Art) (2019)**

*Terræn: Veje ind i samtidskunsten*. Eds. Camma Juel Jepsen Rasmus Kjærboe, Sine Krogh and Martin Søberg. Aarhus University Press, 2019. (The original version in Danish is included here. The English translation of the text is included in the appendix).

The analysis of the three artworks presented in the article, Niels Pugholm's *Emigrant* (2013), *Miss no. 1,2,3* (2013) by Nanna Lysholt Hansen, and *The Woman, The Orphan and The Tiger* (2010) by Jane Jin Kaisei and Guston Sodin-Kung, focuses on the relationship between artworks and listeners – as a dialogue where meaning arises. It is in the artwork's dialogue with its listeners that the voice as a marker of identity is challenged. The technological mediations and manipulations of the vocal expressions in the artworks enable us as listeners to hear the voices as fundamentally paradoxical – as something that transcends the bodily boundaries (by which classical subjectivity is defined) in relation to the listener. The artworks all work with the paradoxical and ambiguous nature of vocal expressions as artistic media in order to articulate issues of emigration, gendered oppression, and the silencing of transnational trauma without reference to specific personal stories or identities. As listeners of the artworks, there is an obligation in the dialogue – to listen to the vocal expressions in the artworks as powerful communications of affective meaning.

### **Vocalizing Dystopian and Utopian Impulses (in press, 2020)**

*The Bloomsbury Handbook of Sound Art*. Eds. Sanne Krogh Groth and Holger Schulze. Bloomsbury Press, 2020 (in press).

In this article, I am concerned with how dystopian and utopian impulses have been vocalized in contemporary artworks as affective vocal expressions. The artworks are

approached through a performative listening, like in *THE END & ROBOT* and *Stemmernes Politik i Samtidskunsten*, and the analysis is grounded in the relationship between artwork and listener. I argue that in the three different artworks discussed, the voices work performatively by the means of technological mediation, manipulation and distribution. The technological stressing of specific elements of the material qualities in the voices, evoke experiences in relation to the audience-listeners, which goes beyond what the voices semantically enunciate. The three different artworks presented in the article, Wagechi Mutu's *The End of Eating Everything* (2013), *Transgressions* (2001) by Nalini Malani and lastly Nanna Lysholt Hansen's *Dear Daughter/Anatomy of the Chthulucene #6* (2018), all work with affective expressions complicating voices as markers of fixed singular identities. Instead, these works point to vocal expressions as something that can momentary become affective places where issues are heard as structural phenomena and could be imagined otherwise in relation to the audience-listeners associations, interpretations and imaginaries.

# Artikel

ROBOT & THE END

# ROBOT & THE END

## A Comparative Critical Reading of the Staging of Synthesized Voices in Digital Media Performances

By Stina Hasse Jørgensen

### Introduction

Today synthesized voices speak to people living in the rapidly developing smart cities in the Western world from technologies such as mobile phones, computers, and social robots, to name just a few. The increasing presence of synthesized voice technologies is also of interest for contemporary artists and the use of synthesized voices as an artistic medium is becoming more and more widespread in media performances, computer games, and new media artworks. It is therefore relevant to discuss how the representation of identities takes place through the use of synthesized voices in the arts. After all, the way we view and identify with “the bodies we see, whether in representation or in real time and ‘live’” informs the way we experience others and ourselves, as art historian Amelia Jones writes in her book *Seeing Differently* (Jones, p. xxi). Furthermore, the representation of bodies in media performances can be said to play a role in the ways we come to define ourselves and others. As media scholar Robin Sloan states: “the media stories that we experience can have significant impact on how we define ourselves as individuals and how we understand others” (Sloan, p. 53).

In this article, I will discuss how representations of gender identity are performed and staged using synthesized voices as an artistic medium in digital media performances. I will critically reflect upon this in my exploration of the 3D animated vocaloid performance *THE END* (2012) and the robotic multimedia dance performance *ROBOT* (2013). In my comparative critical reading of the two media performances I argue that they demonstrate two different ways of using synthesized voices as an artistic medium in digital media performances. More specifically, I will reflect upon the relationship between the sonic body and the visual body in the audio-visual staging and representation of gender identities in digital media performances.

Performance studies researcher Robin Nelson writes in *Mapping Intermediality in Performance* that “the digital doubling of bodies, virtual bodies, robots and cyborgs have entered the intermedial stage, if not to displace humans, then most assuredly to engage with them and question some of their most fundamental assumptions” (Nelson in Bay-Cheng et al., p. 23). Here the use and staging of synthesized voices could promote a proliferation and multiplication of categories, and expand “the range of alternatives, trading duality for multiplicity,” as feminist scholar Mimmi Marinucci writes (Marinucci, p. 47), complicating the fixed visible bodies, and troubling the idea of identity as something static that can be congealed into fixed binaries. In the following I will discuss the audio-visual staging in *THE END* and then in *ROBOT*, reflecting upon the different strategies the two media performances present. I will be concerned with questions inspired by the points made by Nelson and Marinucci, particularly the question of whether digital sonic and visual bodies in media performances have a potential to disrupt the notions of the heteronormative gender binary system.

My comparative critical reading of the media performances *THE END* and *ROBOT* is founded in an understanding of situated knowledge. In her essay “Situated Knowledges: The Science Question

in Feminism and the Privilege of Partial Perspective”, feminist scholar and science & technology theorist Donna Haraway describes situated knowledge as challenging the myths of traditional disembodied objectivity as something that is in fact always already partial, embodied, and specific. Haraway argues that one way of drawing attention to the politics of knowledge production as something that is not disembodied or universal is to make clear that vision is embodied and located. So-called disembodied objectivity, Haraway argues, is already partial, embodied, and specific. The discussion of the two media performances presented in the following is informed by my experiences of being brought up as a white person in the capital (Copenhagen) of a small country (Denmark) and as female assigned at birth. It is also grounded in the located vision defined, among other things, as coming from a situatedness in a Western academic environment where I, as a PhD student at a Department for Arts and Cultural Studies, have been trained mainly in Western art history, sound art, and new media performance with a special interest for artistic experiments with voice technologies.

## THE END

A helium sounding synthesized voice sings “ahhh” in a long tone, weaved together with a rhythmic pulsation of violins. The voice comes from the speakers of a 10.2 surround sound system. It oscillates around the performance space like an echo signal, swirling around me and the other audience members. “Maybe I was dreaming just now as if I was just breathing,” the female voice sings in a staccato rhythm. It comes from the pop star Hatsune Miku, who is the center of attention in the 3D animated vocaloid opera called *THE END*, composed by the acclaimed composer and DJ Keiichiro Shibuya, sound designed by sound artist evala, and 3D animated by art director YKBX. I experienced *THE END* on a summer night in August at Musikhuset Aarhus for the opening of Aarhus Festuge 2016.

*THE END* is a Japanese production, and its special relation to Japanese pop cultural icon Hatsune Miku, who is the lead character in the 3D animated manga-like story with electronic experimental music, attracts many different fans: cosplayers, cartoon fans, anime crowds as well as people interested in experimental music (Jørgensen, 2016a).<sup>1</sup>

*THE END* is a sumptuous feast of technique and aural as well as visual effects. The speakers are placed around the audience in the concert hall and at the front, four large screens form a square around Shibuya, who stands in a yellow hoodie on the stage, surrounded by a smaller screen reminiscent of a DJ booth.

In the performance, a fragmented exploration of the possible disappearance of Hatsune Miku plays out. The narrative unfolds on the large screens on stage in the form of symbolic representations of the feelings the music in the concert hall creates. Throughout the performance, we follow Hatsune Miku’s mental journey and recognition process – from not knowing what it means to exist to understanding that she herself can disappear (like humans do when they die). The display of the end or death of Hatsune Miku has led to many protests from Hatsune Miku’s fans, but despite the protests *THE END* has been touring from Tokyo and Paris to Aarhus since 2012 (Jørgensen, 2016b).

1) There is much more to be written about the situated experience of a Japanese production in a Western context. The discussion of the cultural aspects of *THE END* and how gender performativity might be performed differently in a non-Western context is however not unfolded in this article and the argument presented here leaves room for critical readings by scholars working with intersectionality and postcolonial studies.



*The 3D animated Hatsune Miku in THE END, data.tokyogirlsupdate (2015).*

In Japanese, Hatsune Miku means “the first sound from the future” and first and foremost Hatsune Miku is a sound, a vocaloid. A vocaloid is a voice-based synthesizer technology developed by the music corporation Yamaha and used by media companies like Crypton Future Media, the company behind the release of the Hatsune Miku vocaloid in 2007. Hatsune Miku’s synthesized voice is made from large amounts of vocal recordings by the voice actor Fujita Saki. The digital voice trained on Saki’s voice can be manipulated in the vocaloid software, allowing musicians and producers to create vocal music digitally by typing in lyrics and drawing melodies. They can also automate pitch and singing rate and use effects such as gender factor, growl, and breathiness in the program.<sup>2</sup>

Hatsune Miku was at first codenamed CV01, but was, courtesy of the manga artist Kei, given a visual identity as a 16-year-old manga-like girl with long legs, big eyes, very long turquoise hair in pigtails, often dressed in a miniskirt. This visual depiction of the vocaloid has been further developed through manga drawings, music videos, and anime made by thousands of Hatsune Miku fans worldwide (Jørgensen, Vitting-Seerup, and Wallevik, 2017).

Back in Musikhuset Aarhus concert hall Hatsune Miku appears on the four large screens on stage. ‘She’ is lying down; her long blue hair floating above her head in what could seem like dark water, with two huge cartoon-drawn eyes looking at ‘her’ from behind. ‘She’ looks very

2) In *Synthesis of the Singing Voice by Performance Sampling and Spectral Models* (2007) Jordi Bonada and Xavier Serra write more about how to create singing voice synthesis using audio recordings from voice actors.



fragile as ‘she’ starts to sing with an airy, bright “little-girl voice” as Birgitte Rahbek writes in her review of the performance.<sup>3</sup>

### Gender performativity

In the article “Performativity, Precarity and Sexual Politics” philosopher and queer theorist Judith Butler states that gender performativity is a practice through which gender is constructed through normative constraints and through relations of power: “that gender is performative is to say that it is a certain kind of enactment; the ‘appearance’ of gender is often mistaken as a sign of its internal or inherent truth” (Butler, 2009, p. i). Butler also writes about the notion of gender performativity in *Gender Trouble* (1990) and in *Bodies That Matter* (1993). Here Butler argues that the normative understanding of bodies prompts a performance of gender within a strict binary frame, as a gender binary. The gender binary is the classification of sex and gender into two distinct, opposite and disconnected forms of masculine and feminine. For Butler, the gender binary is part of a regulatory practice “whose regulatory force is made clear as a kind of productive power, the power to produce – demarcate, circulate, differentiate – the bodies it controls” (Butler, 1993, p. xii). Butler examines how the power of heterosexual hegemony shapes normative understandings of bodies, sex, and gender as essential and immutable. Her point is that the regulatory norm of the gender binary is not a given, but is rather constituted by gender performativity, through acts that reiterate the binary of



*Hatsune Miku and catwalk models in clothes designed by the fashion brand Louis Vuitton (Next Nature).*

male and female. She explains gender performativity as something that “must be understood not as a singular or deliberate ‘act’, but, rather, as the reiterative and citational practice by which discourse

3) My translation of “lillepigestemme” (Rahbek, 2016).

produces the effects that it names” (Butler, 1993, p. xii). Butler explains that these normative constraints produce a regulation of which bodies come to matter as socially intelligible and visible subjects, and which bodies are produced as unintelligible and invisible.

### The gender performativity of Hatsune Miku

In *THE END*, Hatsune Miku is depicted with a miniskirt in accordance with previous widespread fan character productions, but here the luxury brand Louis Vuitton has designed the miniskirt. This visual design of Hatsune Miku stages the vocaloid as a digital femininity. This visual depiction of the vocaloid in a miniskirt, with long hair and small waist, supports the auditory experience of Hatsune Miku’s ‘little-girl voice’ as a female audio-visual body, a human sociality.<sup>4</sup> The staging of Hatsune Miku’s sonic and visual body in *THE END* can be argued to play with the notion of presence and human sociality.

Media scholar Thao Phan argues that gendering of synthesized voices, for instance in smart assistants, within the gender binary makes the synthesized voices appear as a “believable performance of human sociality [...] diverting attention away from the act of mediation” in order to design “the illusion of immediacy by which the subject is perfectly seduced by the medium, and in this seduction, indulges in the fantasy that there is no medium at all” (Sloan, p. 28)). The gendering of technologies can create an illusion of immediacy and presence through the design of gendered voices that reiterate the Western binary system. In other words, through a gender performativity constituting the visible subjects of male and female. Maybe the staging of Hatsune Miku plays on the same performance of human sociality as experienced in voice assistants?

Media artist and theorist Aneta Stojnic writes that the increasing personalization of digital devices framed as “smart, intuitive, friendly, responsive, personal, sophisticated” makes them seem human-like, stressing that “all of these terms both attribute and delegate human characteristics to the technologies that become our anthropomorphic companions and our self-extensions” (Stojnic, p. 75). The digital devices are personalized and created as human sociality through the design of, among other things, synthesized voices with personality traits and human paralinguistic characteristics such as gender, age and accent (Nakano, et al.; Baird, et al.).<sup>5</sup> Sound theorist and media philosopher Frances Dyson writes in her book *Tone of Our Times* about this development of synthesized voices: “The quantification of vocal tone, the synthesizing of the voice, and the development of human-computer interfaces based on human-machine conversation, create agents that provide the ‘who’ in this post-theological era” (Dyson, p. 69). The ‘who’ might be agent-based interaction devices such as smart assistants communicating with human users through synthesized voices, as designer Bert Brautigam writes: “You call it Siri, Alexa, or Cortana [...] It defines itself as female or male. The sound of a voice assistant imitates human sound and intonation” (Brautigam, u.p.). The development of synthesized voices to become increasingly human-like has been discussed in connection with what is known as the personification debate (Harris). This debate treats synthesized voices not just as neutral technologies that can be optimized in terms of engineering and programming, but as technologies that are designed within specific cultural understandings of identity construction (Benyon; Faber; Nass & Brave; Robertson).

4) When I write about Hatsune Miku as a ‘female’ character, I do not intend to imply that ‘female’ should be understood as a unified category. On the contrary it is a constructed and complex category highly contested by theorists working intersectionally (e.g. hooks, 1984).

5) For example in IBM~ Bluemix~ you alter such human-like features like “pitch, pitch range, glottal tension, breathiness, rate, and timbre of spoken text” to the synthetic voice (IBM Bluemix 2017).

Just like Siri, Alexa, and Cortana, the synthesized voice of Hatsune Miku is staged as a 'who' – a human sociality with a socially visible and audible body – a female within the gender binary system. The gender representations created in *THE END*, as well as in the smart assistants, demonstrate citational practices that reiterate the normative constraints of the binary system in order to make a "believable performance of human sociality" (Phan). The staging of Hatsune Miku's gendered digital anime body and synthesized 'little-girl voice' constitutes the reiterative and citational practice Butler describes as gender performativity in the context of digital media performance. This citational practice created by the audio-visual presentation of Hatsune Miku can further be argued to fixate 'her' synthesized audio-visual body as a fetish object.

### Fetishizing the audio-visual body of Hatsune Miku

In "Dis-Embodying the Female Voice" critical theorist Kaya Silverman writes about voice-over, arguing that the female voice is always brought back to the female body in cinema – as opposed to the male voice-over, which is often detached from the body, creating an experience of the omnipresent and powerful subject.<sup>6</sup> In *THE END*, Hatsune Miku's voice is never disembodied in the sense that it is always heard as connected to a visual depiction of a 3D animated female body. As such, the visual body and the sonic body of Hatsune Miku are staged in a way that emphasizes the embodiment of the voice in connection with a gender stereotype of the female in the Western gender binary. Following Amelia Jones' reading of the female nude in Western art history, I will argue that Hatsune Miku's audio-visual body is inscribed into the long tradition of the fetish of the 'idealism' of the female form – as it can be seen in e.g. iconic art historical pieces such as *Olympia* (1856) by Édouard Manet, or Alexandre Cabanel's *Birth of Venus* (1863). Here the female nude "must not have any actual genitals: her sex must be erased in order for her body as a whole to function as fetish. The desired female body must, paradoxically, have no orifice, no actual sex" (Jones, p. 65). In *THE END*, Hatsune Miku is not presented with actual genitals or explicit sexual vocal utterings, but appears like an innocent young sexy girl with a miniskirt, big eyes and a 'little-girl voice'. In this way, the presentation of Hatsune Miku can be understood in relation to the aesthetics of the female nude in Western art history and artworks such as Cabanel's *Birth of Venus*. Jones writes that in this particular painting, "the woman, portrayed by the man, is 'deceitful' in her exquisite fleshy offering, and apparently inherently sexually available for heterosexual male gazing." (Jones, p. 65). The vocaloid, Hatsune Miku, can be said to be a female sculpted, quite literally, by a man – namely Shibuya as composer, fan and DJ on stage. Here Hatsune Miku is presented as an aesthetic beauty just as the female (nude) body is presented as a trope of aesthetic beauty in the history of Western art. This is an aesthetic that, following Jones, can be understood as:

*A container to enframe and control the threat of the unbridled female sexuality [...] as a strategic mode of discourse that operates to cohere the male subject, always anxious about the perceived power of female sexuality and social access [...] most often operated in the past through structures of fetishism* (Jones, p. 65).

Just as the control of female sexuality might have been enacted through structures of fetishism throughout Western art history, I will argue that these structures of fetishism are still in play in *THE END*. The staging of Hatsune Miku's voice as dependent on the male composer's desire to

6) Silverman's focus on the female body has been critiqued, arguing that this focus prevents Silverman from seeing a female subject (Fèvre-Berthelot; Sjogren).

listen and play with the voice creates Hatsune Miku as a fetish object. Hatsune Miku is ideal in her feminine desirability. Feminist film theorist Laura Mulvey writes: “Women are simply the scenery onto which men project their narcissistic fantasies” (Mulvey, p. 13). Fetishism functions through a system of binaries, Jones writes, referring here to identity-related discourses on fetishism and especially to psychoanalyst Sigmund Freud’s model of “how objectification occurs in the self/other relation,” where the other can be seen as a “projection of the desires of the empowered self” (Jones, p. 63). In *THE END*, the male composer, fan, and DJ Shibuya projects onto Hatsune Miku, as a representation of the female, his “narcissistic fantasies,” turning her body into a phallic substitute, wholly sexualized and available for his pleasure while not possessing her own sexual identity and body parts (Jones).

Throughout most of *THE END*, Shibuya can be seen standing on stage behind the DJ booth as if he were controlling the voice of Hatsune Miku and determining ‘her’ presence. In a way, Shibuya mimics the role of the Hatsune Miku fan, without which the vocaloid would not exist. Hatsune Miku is to a large extent a crowd-sourced phenomenon and there are more than 100,000 fan-made songs using the voice of the vocaloid to express their emotions and desires. The songs are shared and circulated by fans on social media platforms such as YouTube and Niconico (the Japanese equivalent to YouTube), and the number of songs is still growing. Fans, in this context, should be understood as anyone from bedroom musicians to professional producers and composers such as Shibuya. In *THE END*, Shibuya’s choice to leave the DJ booth could spell the end of Hatsune Miku’s singing. This is what *THE END* is about; the possible end of Hatsune Miku’s voice, abandoned by ‘her’ fans, who no longer desire ‘her’. Hatsune Miku’s audio-visual body is a desired object, a fetish object produced, consumed and shared by fans. In *THE END*, Hatsune Miku is presented as an erotic and commodified “sculpted object”, an object of exchange between fans (of all genders). Hatsune Miku’s audio-visual voice body is an object sculpted by fans, and a token of exchange that “circulates potentially endlessly across time and space, securing a network of future gazes” (Jones, p. 67). The future gaze, which one could argue is a heterosexual male gaze, also has the privilege of seeing without being seen. In *THE END*, this heterosexual male gaze is turned into a fetishizing or listening, where the man or fan, manifested by Shibuya as a male composer and fan, has the privilege of playing and listening to Hatsune Miku’s synthesized voice without his own voice being heard.<sup>7</sup> In this staging of Shibuya as the male counterpart and orchestrator of Hatsune Miku’s feminine appearance and song, Hatsune Miku is inscribed in structures of fetishism as a gender stereotype and a fixed sexualized voice fully assimilated into the desire of others.

### Staging a subversion of gender stereotypes

The staging of the synthesized voice technology that creates Hatsune Miku as a gender stereotype and fetish object can also be used to question gender binary stereotypes. Synthesized voices and digital bodies in media performances have the potential to disrupt the heteronormative gender binary system, complicating the understanding of immediacy as something that can only be achieved by a performance of human sociality through a citational practice reiterating the visible bodies in the gender binary system. Theatre and performance studies scholar Sarah Bay-Cheng

7) It is a difficult task to address the essentialist and dualistic gender politics in the experience of Hatsune Miku’s synthesized voice and visual staging in *THE END* without running the risk of undermining the critique by operating with other dualisms between ‘them’ (the male gaze) and ‘us’ (the feminist) and thereby reintroducing the essentialism being addressed. The notion of ‘us’ and ‘them’ are not global categories and can be parted up into other formations as it is argued for instance in intersectional thinking (hooks).

argues in her article “Virtual Realisms: Dramatic Forays into the Future” that there are many media performances and productions “that take up the questions of virtuality, technological dependence, and digitally transformed bodies” (Bay-Cheng, 2015, p. 689). Bay-Cheng writes that new media offer new possibilities for dramatic theatre to challenge the normative notions of e.g. gender identity (Bay-Cheng, 2015).

An example of this can be found in another media performance using synthesized voices, the multimedia dance performance *ROBOT*. Here seven 58 cm tall humanoid NAO robots are dancing side by side with human dancers.<sup>8</sup> Choreographer Blanca Li, who has collaborated with Pedro Almodóvar, Beyoncé and Michel Gondry, is behind the production. *ROBOT* is a carnivalesque dystopian depiction of humans’ future life in coexistence with robots. The NAO robots have different personalities impersonating human characteristics and emotions, while the dancers mimic the somewhat rigid movements of the robots. In the show, the robots become more like humans. *ROBOT* brings to mind Czech author Karel Capek’s historical play *Rossum’s Universal Robots* from 1920, in which robots display human faults and vulnerabilities. Since *ROBOT* premiered in 2013 it has been performed in more than 100 theatres worldwide, from New York to Braga. I experienced the show in Teatro Circo in Braga, Portugal in the summer of 2015.

When I saw *ROBOT* there was one particular scene that caught my attention: the scene with the “clichéd flirtations” as Zoë Anderson writes in her review of the show. Here I heard the NAO robot’s default synthesized voice for the first and only time in the media performance. The default ‘boyish’ voice was used when the little robot, while performing calisthenics as a tiny version of a bodybuilder, asked the female dancer, lying on the floor next to it: “Don’t you think I’m strong?” making the audience in Teatro Circo laugh, including me. Nuance, the company behind the NAO robot’s default synthesized voice, calls it ‘Kenny’ and describes Kenny as a “he” that will match the NAO robot with a “custom text-to-speech voice specifically designed to match his personality” (Nuance, 2013).<sup>9</sup> This framing of the voice fits with the depiction of the NAO robot as a ‘he’ on stage and in the press material for *ROBOT* (Li). The French male names given to the different NAO robot characters in the show (Sacha, Pierre, Jean, Alex, Lou, Dominique, and Ange) can also be said to promote an interpretation of the robots as masculine, reinforcing the normative constraints of the gender binary (Ellen Jacobs Associates, 2015).

At first, I thought that the gender performativity by the little robot was an example of how digital technologies are staged in media performances as something reiterating the regulatory norms of the female-male dichotomy in a Western gender binary system. I thought that the NAO robot’s performance of a stereotypical ideal of a strong male body in the flirting scene with the female dancer was staging the robot’s male audio-visual body as a fetish object, equivalent to how I have argued that Hatsune Miku is staged as a stereotypical ideal of a female body, as a fetish object or an other desired by the subject in *THE END*. Indeed the clichéd flirtations between the NAO robot and the female dancer in *ROBOT* can be argued to rely on a system of binaries, the gender binary and the binary relation between self and other. However, after reflecting more on the staging of

8) The NAO robots are all-purpose social robots used in service industry, education as well as entertainment (Jørgensen and Tafdrup, 2017).

9) NAO’s plastic robot body first got the ‘Kenny’ voice by Nuance in 2011. In connection with the launch of the collaboration between the then Aldebaran Robotics and Nuance, Steve Chambers, executive vice president in Nuance, talks about the ‘Kenny’ synthesized voice as something that should be perceived of as human-like; “By working with Aldebaran, we’re creating unique and compelling possibilities in the space of robotics where people can connect with NAO as if they were connecting with another human being – and that’s simply powerful” (Rigg, 2013).



the flirtation scene, it occurred to me that the reason why the audience in Teatro Circo laughed was because the little robot's question to the female dancer, which might play with referenced stereotypes of the ideal male body as well as the staging of robots as strong, seemed to fail bravely. It is, after all, ironic and a little bit silly that a tiny (58 cm tall) robot with a very high-pitched voice (as a young boy) that looks more like a cute toy than anything else (with a plastic body and clumsy movements), would ask someone that appears both much larger and stronger to adore the strength of its little body.



*The “clichéé flirtations” between robot and human, Vaison Danses (2014).*

### **The absurdity of the stereotype**

The staged absurdity of the situation lies in its questioning of the robot as something performing human sociality. Here the play between the sonic body and the visual body is crucial. The synthesized voice Kenny and NAO's plastic robot body subvert the experience of the audio-visual robot body as something that can perform human sociality as a male stereotype. This is the opposite of the experience created in the staging of Hatsune Miku in *THE END*. In *ROBOT* the little robot can be understood as an object of attention for the female gaze on stage as well as the audience's gaze. The robot as an object and as clumsy technology is emphasized in its absurd attempt to perform a masculine stereotype and to perform as a visible subject. In this way, the whole scene operates within dichotomies of the binary system, of male-female and subject-other, yet subverts the gender performativity within the frame of the binary system by questioning our, the audience's, notion of gender stereotypes and robot stereotypes.

The subversion of gender stereotypes is also explored at another point during the show. Here, the same NAO robot is dressed up in a purple sequined dress and a pink boa, dancing, and singing to the song *Bésame Mucho*. This time, however, the song is not performed by NAO's default synthesized voice Kenny, but is a playback recording sung by a female singer. In this way the robot's body is connected to two different sonic bodies, a male and a female, underlining the arbitrariness (and absurdity) of the gender performativity in humanoid robot technology (and technologies designed to be human-like).

The absurdity of gendering the robot technology within a Western binary system as a performance of human sociality is further stressed by the use of playback technology. If this scene is an attempt to create an experience of a performance of human sociality, it fails, since the act of mediation, the playback technology, is obvious to the audience. The expressional vocabulary and gestures of the little plastic robot seem disconnected from the playback voice and do not match the emotional intensity of the playback song. Again, the robot is staged as an object in its absurd attempt to perform a stereotype and to perform as a visible subject.

This subversive play with stereotypes is something that Blanca Li, the choreographer of *ROBOT*, is known for. Her play with gender identities can be experienced in the music video *Around the World* by the electronic music duo Daft Punk, where she humorously had dancers perform different kinds of stereotyped bodies: skeletons, zombies, and robots.

## Conclusion

In this article, I have discussed how representations of gender identity are performed and staged using synthesized voices as an artistic medium. In a comparative critical reading of two digital media performances, *THE END* and *ROBOT*, I have pointed out two different strategies for using synthesized voices as artistic medium, especially in relation to the interconnectedness of the synthesized voices and their visual framing (the sonic body and the visual body) in the audio-visual staging and representation of gender identities in digital media performances.

I have argued that the staging of Hatsune Miku's synthesized voice in the vocaloid performance *THE END* can be said to perform and reiterate a normative binary system, creating the audio-visual body as a fetish object assimilated to the desire of others. In this way, *THE END* can be argued to display the design and performance of a synthesized voice operating within binaries such as male-female and subject-other. As such, it somehow seems to ignore the critique of the binary system made by feminists, queer theorists, and poststructuralists throughout the last 35 years.<sup>10</sup>

The staging of the synthesized voice Kenny in the media dance performance *ROBOT* also arguably displays a design and performance of the synthesized voice as reiterating the regulatory norms of the Western gender binary and the binary of the subject-object. Contrary to the staging of Hatsune Miku in *THE END*, however, the staging of Kenny and the NAO robot in *ROBOT* seems to disrupt the notions of the heteronormative gender binary system. By staging the irony of the gender performativity of a robot with a little plastic body, and by further complicating the understanding of the interconnectedness between the sonic and the visual body, giving the same plastic robot a male and a female voice, the citational practice of reiterating the visible bodies in the binary system is challenged. *ROBOT* is a media performance that troubles the idea of identities

10) The gender stereotypes performed with the use of synthesized voices in both *THE END* and *ROBOT* can also be experienced in technologies with synthesized voices, most widely known in from the smart assistants Siri by Apple, Alexa by Amazon, Cortana by IBM, that all operate with a categorized male or female voice. Another example is the IBM Bluemix Catalogue with thirteen voices categorized as either male or female (Baird 2018; IBM Bluemix 2017).

and gender representations as fixed through a subversion of the staging of gender binaries and stereotypes.

My comparative critical reading of the media performances presented here has demonstrated how the discussion of synthesized voices as an artistic medium should take into account the interplay between the auditive performance, the visual framing, and the situated context. As an artistic medium, the synthesized voice can be said to operate with a gender performativity that creates stereotypical gender representations. However, the staging of synthesized voices in media performances matters, since it is possible to use the connection between the synthesized voices and their visual framing (the sonic body and the visual body) subversively. Future media performances with synthesized voices might even challenge the normative notions of gender identity and promote a multiplication of representational categories and bodies of difference.

---

### Stina Hasse Jørgensen

is a PhD student at the the Department of Arts and Cultural Studies at the University of Copenhagen. In her research, Jørgensen has focused on topics such as sound art, media art, feminism, gender theory, vocal theory, interaction design, performativity and performance art. She has written articles and contributions to anthologies, online magazines like *Kunsten.nu* and *Seismograf* as well as for journals such as *Digital Creativity; Transformations; Body, Space & Technology Journal*; and *Cultural Analysis Journal*.

---

### References

- Anderson, Z., 2017. *Robot, Barbican Theatre, London, review: The audience gasps in dismay when it falls over, and laughs fondly when it wiggles its fingers, asking to be lifted.* [online] Available at: <http://www.independent.co.uk/news/robot-barbican-theatre-london-review-blanca-li-dance-a7601601.html> [Accessed: April 19th 2018].
- Baird, A., Jørgensen, S.H., Parada-Cabaleiro, E., Cummins, N., Hantke, S., Schuller, B., 2018. The Perception of Vocal Traits in Synthesized Voices: Age, Gender, and Human-Likeness. *JAES Journal of the Audio Engineering Society*, vol. 66(4), pp. 277-285.
- Bay-Cheng, S., Kattenbelt, C., Lavender, A., and Nelson, R., 2010. *Mapping Intermediality in Performance*. Amsterdam: Amsterdam University Press.
- Bay-Cheng, S., 2015. Virtual Realisms: Dramatic Forays into the Future. *Theatre Journal*, 67, pp. 687-698.
- Benyon, D., 2014. *Designing Interactive Systems: A comprehensive guide to HCI, UX and interaction design, 3/E*. Harlow: Pearson.
- Bonada, J., and Serra, X., 2007. Synthesis of the Singing Voice by Performance Sampling and Spectral Models. *IEEE Signal Processing Magazine*, 24(2), pp. 67 - 79.
- Brautigam, B., 2017. *The New Skeuomorphism is in Your Voice Assistant.* [online] Available at: <https://uxdesign.cc/the-new-skeuomorphism-is-in-your-voice-assistant-3b14a6553a0e> [Accessed: March 16th 2018].
- Bremers, A., 2015. "The End": a Virtual Opera Singer, Millions of Fans and the Meaning of Death." *Next Nature Network*, [image online] Available at: <http://nextnature.net/2015/09/end-opera-players-virtual/> [Accessed: April 28th 2018].
- Butler, J., 1993. *Bodies That Matter*. New York: Routledge.



Butler, J., 2009. Performativity, Precarity and Sexual Politics. *AIBR. Revista de Antropologia Iberoamericana*, 4(3), pp. i-xiii.

Data.tokyogirl., 2015. *Miku Hatsune's Vocaloid Opera "THE END" to Be Performed in Shanghai Next!* [image online] Available at: <https://tokyogirlsupdate.com/miku-the-end-shanghai-20150751579.html> [Accessed: May 26th 2018].

Dyson, F., 2014. *The Tone of Our Times: Sound, Sense, Economy, and Ecology*. Massachusetts: The MIT Press.

Ellen Jacobs Associates., 2015. *From Paris with Robots*. [image online] Available at: [http://www.ejassociates.org/press\\_releases/from-paris-with-robots/](http://www.ejassociates.org/press_releases/from-paris-with-robots/) [Accessed: May 18th 2018].

Faber, L. W., 2013. *From Star Trek to Siri: (Dis)Embodied Gender and the Acousmatic Computer in Science Fiction Film and Television*. Ph.D. Southern Illinois University Carbondale.

Fèvre-Berthelot, A. L., 2013. Audio-Visual: Disembodied Voices in Theory. *InMedia*, [online] Available at: <http://journals.openedition.org/inmedia/697> [Accessed: May 20th 2018].

Haraway, D., 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14 (3), pp. 575-599.

Harris, R. A., 2005. *Voice Interaction Design: Crafting the New Conversational Speech System*. San Francisco: Morgan Kaufmann Publishers/Elsevier.

hooks, bell. 1984. *Feminist Theory: from margin to center*. New York: Routledge.

IBMBluemix, 2017. *Using SSML*. [online] Available at: <https://console.bluemix.net/docs/services/text-to-speech/SSML.html#ssml> [Accessed: April 25th 2018].

Jones, A., 2012. *Seeing Differently*. New York: Routledge.

Jørgensen, S. H., 2016a. *Popstjerne af lys, lyd og software*. [online] Available at: <https://kunsten.nu/journal/popstjerne-lys-lyd-software/> [Accessed: April 15th 2018].

Jørgensen, S. H., 2016b. *THE END: En Teknisk Operaoplevelse*. [online] Available at: <http://seismograf.org/the-end-en-teknisk-operaoplevelse> [Accessed: May 26th 2018].

Jørgensen, S. H., Tafdrup, O., 2017. Technological Fantasies of Nao - Remarks about Alterity Relations. *Transformations*, 29, pp. 88-103.

Jørgensen, S. H., Vitting-Seerup, S., and Wallevik, K., 2017. Hatsune Miku : An uncertain image. *Digital Creativity*, 28(4), pp. 318 – 331.

Li, B., 2015. *Robotic pop ballet for all future generations*. [online] Available at: [blancali.com/en/download/Robot2015-english.pdf/pdf\\_120\\_file\\_fr.pdf](http://blancali.com/en/download/Robot2015-english.pdf/pdf_120_file_fr.pdf) [Accessed: March 18th 2018].

Marinucci, M., 2016. *Feminism is Queer: The intimate connection between queer and feminist theory*. London and New York: Zed Books.

Mulvey, L., 1989. *Visual and Other Pleasures*. New York: Palgrave Macmillan.

Nakano, Y., Neff, M., Paiva, A., Walker, M. ed., 2012. *Intelligent Virtual Assistants*. Berlin: Springer-Verlag.

Nass, C. and Brave, S., 2005. *Wired for Speech: How Voice Activated and Advances the Human- Computer Relationship*. Massachusetts: The MIT Press.

Nuance, 2013. *Aldebaran robotics and Nuance revolutionize human-machine interaction*. [online] Available at: <http://whatsnext.nuance.com/connected-living/aldebaran-robots-nuance-revolutionize-human-machine-interaction/> [Accessed: May 17th 2018].

- Phan, T., 2017. The Materiality of the Digital and the Gendered Voice of Siri. *Transformations*, 29, pp. 23-33.
- Rahbek, B., 2016. *Hun er 16 år, har millioner af fans - og findes ikke*. [online] Available at: <https://www.b.dk/nationalt/hun-er-16-aar-har-millioner-af-fans-og-findes-ikke> [Accessed: February 18th 2018].
- Rigg, J., 2013. *Nao robot to become even more of a chatterbox with new software (video)*. [online] Available at: <https://www.engadget.com/2013/10/30/nao-robot-new-nuance-voice-software/> [Accessed: May 26th 2018].
- Robertson, A., 2016. *Google's DeepMind AI Fakes Some of the Most Realistic Human Voices Yet*. [online] Available at: <https://www.theverge.com/2016/9/9/12860866/google-deepmind-wavenet-ai-text-to-speech-synthesis> [Accessed: May 25th 2018].
- Silverman, K., 1984. Dis-Embodying the Female Voice. In: Doane, M. A., Mellencamp, P., Williams, L., eds. 1984. *Re-vision: Essays in Feminist Film Criticism*. Los Angeles: University Publications of America, pp. 131-149.
- Sjogren, B., 2006. *Into the Vortex – Female Voice and Paradox in Film*. Chicago: University of Illinois Press.
- Sloan, Robin J.S., 2015. *Virtual Character Design for Games and Interactive Media*. Boca Raton, FL: CRC Press.
- Stojnic, A., 2015. Digital Anthropomorphism. *Performance Research: A Journal of The Performing Arts*, 20(2), pp. 70-77.
- Vaison Danses., 2014. *Blanca Li - Robot ! - Festival Vaison Danses 2014*. [image online] Available at: <https://www.youtube.com/watch?v=VHCA1eEW7jI> [Accessed: February 26th 2018].



# Stemmernes politik i samtidskunsten

Stina Hasse Jørgensen

Overalt er der stemmer. Stemmer fra andre og fra os selv. Nogle passerer forbi os, imens andre er vedrørende og påtrængende. Klangflader, rytmer og stød mærkes i vores ører, idet stemmerne taler. Forestillinger om intimitet, identitet, autoritet og tilhørsforhold dannes i mellemrummet mellem stemmerne og os, der lytter til dem. Vores indre stemme blandes med de stemmer, vi hører.<sup>59</sup> Stemmen er altid interrelationel. Vi kan ikke sætte os over, under eller uden for stemmerne. Når vi lytter til andres tale, former måden, de taler på, den accent, de har, de ord, de vælger at bruge, vores oplevelse af dem ud fra sociale, kulturelle og politiske normer. På grund af stemmernes væsentlighed og allestedsnærvær optræder de også som medie eller omdrejningspunkt i samtidskunsten. I dette bidrag vil jeg præsentere udvalgte danske samtidskunstværker og diskutere, hvordan de på forskellige måder arbejder med stemmernes politik i en dialog med lytteren. Bidraget er blevet til på baggrund af et længere samarbejde med Jane Jin Kaisen omkring projektet *For More Than One Voice* (2016-), der omhandler flerstemmighedens politik og poetik, samt igennem diskussioner og interviews med kunstnerne Niels Pugholm og Nanna Lysholt Hansen.

## Stemmen som kunstnerisk medie

Stemmeteoretikeren Nori Neumark skriver i introduktionen til bogen *VØICE* fra 2010, at stemmen i dag er kommet tilbage på den internationale teoretiske og kunstneriske agenda. Siden starten af 2000 er et væld af bøger, artikler og kunstværker, som behandler stemmen som et sprogligt, lydligt og kropsligt fænomen, blevet offentliggjort og udstillet (Neumark 2010: xv).<sup>60</sup> Lydteoretikere og kunstnere som Mary

Kelly, Miriama Young, Lawrence Abu Hamdan og Brandon LaBelle har for eksempel skrevet om stemmens særegenhed som kunstnerisk medie. Her er et fokus på, hvordan stemmen kan skabe en dialog mellem værket og lytteren. I bogen *Dialogue – On the Politics of Voice* skriver Kelly i introduktionen, at når vi fokuserer på stemmen i kunsten, er det som at stille skarpt på perspektivet i et maleri eller på en performancekunstners bevægelser og påklædning som væsentlige elementer i vores oplevelse af værkets meningsskabelse (Kelly 2011: 12). Stemmen skaber på samme måde betydning i dialogen mellem værk og beskuer. Beskueren er her også en lytter, der fortolker stemmen i værket, på samme måde som en tilskuer registrerer perspektivet og observerer bevægelser, farver og komposition i værket og forstår dem som væsentlige elementer i kunstens betydningsdannelser. Young skriver i bogen *Singing the Body Electric* fra 2015, at stemmen kan indfange lytteren med dens undvigende kvaliteter: Som lytter kan man blive draget af stemmens materielle, sanselige udtryk og reflektere over dens metaforiske betydninger. Men vi kan aldrig indfange stemmen fuldt ud, skriver hun, for heri ligger dens kunstneriske potentiale (Young 2015: 2). Med artiklen "Aural Contract: Forensic Listening and the Reorganization of the Speaking Subject" fra 2012 gør Hamdan også opmærksom på lytterens rolle i forhold til stemmen. Han ønsker at generere en diskussion om, hvordan vi lytter til stemmer, og hvordan denne lytning hænger sammen med bredere politiske og sociale interesser, hvilket også afspejles i Hamdans værker som for eksempel *The Whole Truth* (2012) eller *Conflicted Phonemes* (2012). Værkerne problematiserer fremkomsten af stemmeanalyseapparater som efterforskningsredskaber inden for retslige instanser. I *Conflicted Phonemes* undersøger Hamdan eksempelvis, hvordan særlige måder at lytte på er forbundet med specifikke sociale og politiske forståelser af det talende subjekt. I værket udforsker Hamdan blandt andet, hvordan stemmeanalyseapparater fordrer kriminaltekniske lyttemetoder, der rubricerer det talende subjekt inden for et ensidigt system af kategorier, som ikke kan rumme stemmens flertydigheder og foranderlighed. I værket *Channels* (2013), skabt af kunstneren Susan Hiller, hører vi beretninger om forskellige menneskers dødsoplevelser. Her sammenkædes stemmerne i en undersøgelse af vores kulturelle oplevelser af,

hvordan vi lytter til stemmer som vidnesbyrd. Stemmen som vidnesbyrd undersøges på en anden måde i kunstneren Arthur Watsons konceptuelle performance *Singing for Dead Singers* (2000), hvor Watson synger ballader på en skotsk dialekt, som kun eksisterer oralt, og som vil forstumme, hvis ikke sangene bliver sunget som en manifestation af denne dialekt og musikalske tradition.

I bogen *Background Noise: Perspectives on Sound Art* fra 2006 skriver LaBelle, at brugen af stemmen i kunsten ofte kan sammenkædes med en politisk og social kunstnerisk praksis. En kunst, der blandt andet behandler spørgsmål om identitet og tilhørsforhold, som forholder sig til, hvordan man konstitueres og konstrueres som subjekt eller objekt, og hvordan dette positioneres og performes i det sociale og politiske spektrum (LaBelle 2006: 101). Det kan siges at gælde for Hamdans, Hillers og Watsons værker, og det kan også siges at gøre sig gældende inden for den danske samtidskunstscene. Man kan for eksempel opleve dette i værker som *The Image of Me* (2012) af Jeanette Ehlers og Patricia Kaersenhout, Hannah Heilmanns tv-kunstværk *Allons-y* (2016) og Kristoffer Ørums *Data-kroppe* (2016). Værkerne iscenesætter på forskellig vis en lytten til stemmen som noget, der kan destabilisere sociale konventioner, normdannelser eller oplevelser af subjektet som fast forankret og entydigt. Her er stemmerne, kroppe og dialogen med lytterne i værket med til at definere, udpege og overskride grænserne og mulighederne for de individuelle tilstedeværelser. I værkerne *Emigrant* (2013) af Niels Pugholm, Nanna Lysholt Hansens *Miss no. 1, 2, 3* (2013) og *The Woman, the Orphan, and the Tiger* (2010) af Jane Jin Kaisen og Guston Sondin-Kung er der også en dialog mellem værk og lytter. Antropologen Amanda Weidman skriver i artiklen "Anthropology and Voice" fra 2014, at det er gennem den måde, hvorpå stemmerne giver resonans i lyttere og offentligheder, at meningen bliver hørbar (Weidman 2014: 45). Meningen i disse værker fremtræder igennem dialogerne med lytteren. Det er dialoger, som komplicerer og stiller spørgsmål til identitetskategorier og tilhørsforhold på måder, der peger på politiske systemer og sociale problematikker, som findes globalt og nationalt. Ved at sætte stemmen som materiel praksis og metafor til forhandling i relationen mellem værk og lytter indkredser og problematiserer værkerne stemmen som karakter, poli-

tisk udsagn og vidnesbyrd. I det følgende vil værkerne blive diskuteret med dette for øje i en bestræbelse på at forstå, hvilken rolle stemmen spiller i den danske samtidskunst i dag.

### Emigrant

"Stemmen er et tegn, ligesom røgen fra en pistol er et tegn på, at den har været affyret," siger kunstneren Niels Pugholm i et interview og fortsætter: "Hvad sker der, når man snakker dansk med tydelig accent? Hvilket billede skaber det hos lytteren? Jeg bruger stemmen som et kunstnerisk virkemiddel, en tegnværdi, der skal fortolkes." (Pugholm 2016)

Stemmen som tegn er et af de elementer, som Niels Pugholm undersøger i sit værk *Emigrant* fra 2013. Værket består af en 13 minutter lang lyd- og videoinstallation og er blandt andet vist til *Sektor 13* på Brandts i 2013 og til *FOKUS 2015* på Nikolaj Kunsthal. I værket ser vi på drivskyer, der langsomt bevæger sig over en lyseblå himmel og hører en stemme sige: "Skyerne ligner hinanden overalt på kloden. Det er de samme, uanset hvorhen du kommer. De hænger bare deroppe, som små, hvide puder på en blå himmel, eller lægger sig over dig som et blødt tæppe. Nogle gange er de helt væk, men du er sikker på, at de kommer tilbage igen. Jeg kan godt lide skyer." Den visuelle billedside knyttes til den auditive side af værket. Skyerne vandrer roligt fra et sted til et andet og ud af vores synsfelt, efterhånden som værket udfolder sig over tid.

Stemmen fortsætter: "Jeg blev født på landet. Mine forældre havde en lille bondegård, og vi var ti børn i familien. Jeg havde fire søstre, som var ældre end mig selv, men jeg var den ældste af drengene." Det er en poetisk, blid mandestemme, som taler dansk med tydelig accent. Han krøller ordene blødt sammen med rullende r'er og bølgender uden stød. Det indikerer, at han har et andet modersmål end dansk, selv om han taler sproget flydende. Vi får derfor, på dette tidspunkt i værket, et indtryk af, at personen, hvis stemme vi lytter til, fortæller om sin egen opvækst på landet med hårdt arbejde fra en tidlig alder. Titlen, *Emigrant*, og accenten understøtter vores sammenkædning af stemmen, vi hører, med historien, der bliver fortalt. Det er en emigrationsfortælling om det at forlade sit hjemland for at søge lyk-



ken og et bedre liv i et andet land, i dette tilfælde Danmark. Vi forstår ham godt. Der er jo dejligt i Danmark. Vores forestilling om, at emigrationshistorien tilhører stemmen, vi lytter til, smuldrer dog, da vi hører, at beretningens ”jeg” bliver ramt af den spanske syge under den store epidemi og lidt senere, en måned efter første verdenskrig, tager af sted til Amerika for at prøve lykken.

Emigrationsfortællingen er inspireret af Niels Pugholms morfars storebror, en etnisk dansker ved navn Christian ”Chris” Pedersen (1902-1997). Stemmen, vi hører oplæse fortællingen i værket, tilhører Georgius Skouros, som emigrerede fra Grækenland til Danmark i 1980’erne.

De hvide drivskyer glider over himlen, mens en fortælling om immigration udfoldes i videoværket *Emigrant* af Niels Pugholm fra 2013.

### Stemmen som tegn

I Pugholms værk er Skouros’ stemme et tegn. Ligesom røgen fra pistolen er et tegn på, at den har været affyret, er måden, Skouros taler på, hans flydende dansk med tydelig udenlandsk accent, et tegn på en særlig befolkningsgruppe, som er indvandret fra et andet land til Danmark. I værkets begyndelse sammenknytter vi som lyttere let Skouros’ vokale praksis med fortællingen om emigration. Vi kommer også naturligt til at forestille os, at værket handler om en særlig befolkningsgruppes immigration til Danmark. I *Emigrant* stiller Pugholm således spørgsmålstejn ved vores lytteoplevelse af Skouros’ vokale praksis som tegn på en specifik identitet, som en metafor for en person, der ikke er født i Danmark, men som er rejst hertil i ønsket om at få et bedre liv. Idet vi finder ud af, at det ikke er Skouros’ egen emigrationsfortælling, bliver vores oplevelse af Skouros’ stemme som tegn på en bestemt identitet punkteret. I løbet af værket opstår der en forhandling mellem værket og os som lyttere, om hvem det er, vi hører tale. Det, at vi aldrig får et visuelt billede af den, der taler, understøtter tvivlen om, hvem der egentlig er bag stemmen og fortællingen. Værket fremstår nu tvetydigt og komplekst. Stemmen, vi lytter til, er sammensmeltet med ordene, oplæsningen af fortællingen i værket, men samtidig står hvert element frem hver for sig. Stemmen har fået en ny betydning som tegnværdi i den simultane sammensmeltning og adskillelse af stemmen som materiel praksis og som metafor. Stemmen, vi hører, forbinder 1910’ernes emigration fra Danmark til USA med immigrationen til Danmark fra 1970’erne og frem til i dag. Fortællingerne er ét sammenhængende hele og er alligevel hver for sig. Værkets visuelle del understøtter dette, idet personerne er auditivt nærværende, men ikke synlige kropsligt. I bogen *The Voice in Cinema* forklarer film- og lydteoretikeren Michel Chion udtrykket *acousmètre* som et filmisk redskab, et greb, hvor en stemmekarakter fremstår uden krop. Stemmen vandrer på skærmens overflade uden at gennemtrænge den. Som beskuer og lytter kan vi kun forestille os og fantasere om stemmens krop. I dette *acousmètre*, denne spænding, dette equilibrium, hvor vi får lov til at høre, men ikke se, ligger der en kraft. Stemmen fremtræder som omnipotent, som allestedsnærværende (Chion 1999: 24). Stemmen i *Emigrant* er et *acousmètre*. Vi ser aldrig hverken Chris eller Georgi-



us, men følger drivskyernes rejse over himlen, som et globalt vejrphænomen. Værket fremstår som et poetisk konglomerat af fortællingen, stemmen og skyerne på himlen. Stemmen er ikke længere en identitetsmarkør for folk, der er immigreret til Danmark. Stemmen er et tegn på emigration som et transnationalt og transhistorisk fænomen, som en personlig, men også almen fortælling på tværs af historien og forskellige kontinenter.

### Miss no. 1, 2, 3

Kunstneren Nanna Lysholt Hansen skabte i 2013 videoværket *Miss no. 1, 2, 3*. Det er blandt andet blevet vist i 2014 på udstillingen *Body Screeners* i NLHspace, men som mange af Lysholt Hansens andre værker er *Miss no. 1, 2, 3* et led i en lang række værker, der arbejder med samme omdrejningspunkt. Allerede i 2009 skabte Lysholt Hansen *Miss*, et videoværk til Forårsudstillingen på Charlottenborg, hvor det engelske ord "miss" blev læst op omkranset af ord begyndende med "mis-". *Miss* er også blevet vist på udstillingen *We Are Grammar* på Pratt Manhattan Gallery i New York i 2011. Senere samme år fulgte en videreudvikling af værket, *Miss no. 1,2,3, live # 1*, hvor tre ordlister relateret til ordet "miss" blev opført som liveperformance i forbindelse med *Afgang 2011* på Kunsthallen Nikolaj. Videoværket *Miss no. 1, 2, 3* bygger videre på de to foregående værker. Her hører vi definitioner af "miss" i dets betydning som verbum og navneord oplæst side om side med ordets synonymmer, som de står ordret i ordbogen.<sup>61</sup> I forsøget på at indkredse ordets betydningsskabelse og sanselige materialitet i værket er Lysholt Hansens arbejde med stemmen som kunstnerisk medie centralt. "Hvilke slags stemmer lytter vi til? Hvad betyder det, om de er kønnede? Og hvad er forskellen på, om der er én eller mange stemmer, der taler?", spørger Nanna Lysholt Hansen og fortsætter: "For mig handler det om at stille spørgsmålene subtilt i værket, på en måde, der åbner op for refleksion hos lytteren." (Lysholt Hansen 2016) Spørgsmålet om stemmen, hvem eller hvad vi lytter til, bliver udfoldet på flere måder i værket i en sanselig leg med publikums position: Hvem er subjekt, og hvem er objekt? Hvem bliver set, og hvem ser? Hvilke stemmer bliver der lyttet til og hvordan?



I *Miss no. 1, 2, 3* (2013) ser vi tre næsten identiske videoklip af Lysholt Hansen. I alle tre ser vi kunstneren afbildet som buste med de samme nøgne skuldre og lyse ansigt med håret hængende ned foran ørerne. Vi ser dem, klonerne, som Lysholt Hansen kalder dem, kigge direkte på os, tavse. Så begynder den midterste klon at tale: "Miss, Missed, Missing, Misses". Stemmen er ulden og dyb. Den lyder hverken som en mand eller kvinde, men som noget, der balancerer midt mellem kønnene. Efter et halvt minut istemmer klonen til venstre med opremsninger af ord med "miss" eller "miss-" foran, for eksempel: "mis-apprehend, mis-apprehension, mis-appropriate, mis-appropriation." Det lyder ikke som den midterste klons stemme. Den er lidt lysere og ligger i krydsfeltet mellem en tenor- og en altstemme. Den sidste klon blander sig i ordstrømmen: "Miss, a young woman; a young lady of 18, fille, girl, missy, young lady, young woman, babe, sister, baby, sometimes used as a term of address for attractive young women." Stemmen er, som de andre kloners, metallisk mørk. Ordene knytter sig til de andres slæbende, taktfaste rytme. Vi kommer i tvivl om, hvorvidt vi kan placere ansigterne, vi ser, inden for et binært kønsklassifikations-system i deres sammenkobling med stemmerne. I brudstykker tager sætningerne form som en mærkværdig flerstemmig samtale mellem de tre kloner. "Miss. Used as a courtesy title before the surname or full

I Nanna Lysholt Hansens videoværk *Miss no. 1, 2, 3* fra 2013 bliver beskueren forført og forundret over stemmerne, der høres.

name of a girl or single woman,” siger klonen i midten, alt imens den sidste klon fortsætter med sine beskrivelser: ”Bimbo, a young woman indulged by rich and powerful older men.” Ordene flyder ind og ud af klonerne, som er de masseproducerende ordmaskiner. De taler ryt-misk, i munden på hinanden, ved siden af hinanden. Ansigter, stemmer og ord blandes sammen: ”Used as a form of polite address for a girl or young woman. A young unmarried woman.” Miss 1, 2 og 3 kigger alvorligt på os med indtrængende øjne. Vi skal se på dem. Vi skal lytte til deres stemmer. Vi skal høre listerne af ord, de opremser, sammen og hver for sig.

### **Stemmernes anfægtelse af ordenes politik**

Lydteoretikeren Douglas Kahn skriver, at teknologi muliggør en situation og en konversation, der ellers ikke ville have fundet sted – en forvandling af materialiteten, en opløsning eller ændring af betydning, i en situation, hvor kunstneren har sat rammerne for publikums deltagelse i mødet med kunsten (Kahn 1999: 3). Lysholt Hansen arbejder med stemmemanipulation, toneleje og rytmik, så vi bliver suget ind af rytmen og lokket af stemmernes tvetydigheder og sensuelle balance mellem kønsstereotype grænser. Deres klangfarver og pitch udfordrer vores oplevelse af, at stemmen kan være en metafor for en identitet, som kan hægtes op på et køn. Værket peger på, hvordan vi ofte sætter lighedstegn mellem stemme, repræsentation og identitet og ignorerer kompleksiteterne i, hvordan en stemme reelt er konstrueret, medieret og hørt. I værket hører vi stemmerne som fundamentalt paradoksale: som fænomener, der konstituerer, destabiliserer og forvrænger vores forestillinger om identitet. Værket rusker i vores forestilling om stemmen som en metafor for en subjektivitet, som en repræsentation af ét talende subjekt med én identitet og subjektivitet. Vi får som lyttere ikke en mulighed for at definere, klassificere og rubricere hver enkel ”miss” som et talende subjekt med én identitet, for de er pitchet, reproduceret og medieret igennem forskellige teknologier. Neumark skriver, at stemmen medieres igennem teknologier såvel som igennem kulturer (Neumark 2010: xviii). Igennem en manipulation af stemmernes pitch peger værket på, hvordan de allerede er medieret af vores kulturelle forestillinger, og hvordan allerede eksisterende

forståelser af stemmen som noget, der tilhører en kønnet krop inden for et binært kønssystem – kvindestemmen tilhører kvindekroppen, og mandestemmen tilhører mandens krop – dominerer vores lytten til stemmen.

Queer-teoretikeren og filosofen Judith Butler skriver om, hvordan navngivning og benævnelse konstituerer et subjekt igennem sproget (Butler 1997: 5). I bogen *Excitable Speech: A Politics of the Performative* spørger hun: Kan vi forestille os en krop, som ikke er blevet tilskrevet nogen sociale definitioner? Hvilken social eksistens har denne krop? Er det ikke først i kaldet og adresseringen, at kroppen træder frem for os? I værket problematiseres tiltalen og navngivning som en måde, hvorpå sproget kategoriserer kroppen. Hvad sker der, hvis kroppen ikke falder inden for de kategorier, som er tilgængelige for os i sproget? Hvis man er mere, både og, og andet end det, som tiltalen dækker over? Kan ordene rumme os som mennesker med flydende seksualitet og køn?

Værkets fremvisning af de sociale strukturer, som klassificerer stemme og krop til at skulle kunne identificeres inden for et binært kønssystem, handler ikke kun om stemmen, men også om sproget. Ordene bruger vi til at tiltale folk forskelligt, alt efter hvilket køn eller hvilken status de har. Igennem værkets destabilisering af stemmen som metafor for en identitet sker der også noget mere. Vi bliver fremmede over for ordene, vi lytter til. De er blevet til objekter, vi kan anskue på ny. Hvorfor findes der forskellige benævnelser af en kvinde baseret på, hvorvidt hun er gift eller ej? Og hvorfor har dette ”miss” synonyme som ”bimbo”, ”sexbomb” og ”mistress”? Hvorfor kædes disse betegnelser af en ugift kvinde sammen? Lysholt Hansens værk får os til at tænke kritisk over de måder, sproget klassificerer og skaber sammenhænge på, når vi taler: For er ”miss” og alle dets associationer ikke en manifestation af politiske kønsbaserede magtstrukturer i sproget?

### **The Woman, The Orphan, and The Tiger**

Den 72 minutter lange, narrativt eksperimenterende film *The Woman, The Orphan, and The Tiger* fra 2010, af kunstnerne Jane Jin Kaisen og Guston Soudin-Kung, skaber en strategisk politisk genealogi mellem



tre generationer af kvinder; de omkring 200.000 kvinder, der blev udsat for militært seksuelt slaveri af det japanske militær mellem første og anden verdenskrig, den anslåede ene million kvinder, der har arbejdet som sexarbejdere omkring amerikanske militærbaser i Sydkorea fra 1950'erne til i dag, og de cirka 200.000 børn, der er blevet adopteret fra Sydkorea til Vesten siden 1950'erne. Siden 2010 er filmen blevet udstillet og screenet over hele verden, blandt andet Videonale13 (Tyskland), Space \*C, Coreana Museum of Art og Jeju Intl. Women's Film Festival (Sydkorea), Yamagata Intl. Documentary Film Festival (Japan), Taiwan Intl. Documentary Film Festival, DePaul Art Museum, New Wight Gallery, Vox Populi Galleri og Korean American Film Festival New York (USA).

Filmen afdækker, hvordan militær og patriarkalsk vold mod kvinder og børn blev central i geopolitiske forhandlinger mellem Sydkorea, USA og Japan, og hvordan denne del af historieskrivningen i høj grad har været fortiet og undertrykt, men til stadighed vækker genklang i samtiden. Dette fremstilles i filmen igennem en kompleks fortælling om traumer og erindringer, som ikke udfoldes lineært kronologisk, men i stedet drives frem i fragmentariske sekvenser med fokus på stemmerne, vi hører. Stemmerne er et kunstnerisk omdrejningspunkt og virkemiddel, som bliver slået an fra filmens begyndelse. Her ser og hører vi, hvordan en ældre kvinde bag et vidnepanel i en retssal rejser sig med bistand fra en yngre kvinde for at træde frem mellem vidnepanelet og dommerstolen. Kvindens hoved er bagoverbøjet, og hendes øjne er næsten lukkede. Hun vender sig mod dommeren og slår sig med sin venstre hånd gentagne gange på hjertet, før hun vender sig tilbage mod den yngre kvinde, der rækker ud for at hjælpe hende. Hun undslår sig og vender sig om mod retssalen. Hendes hånd bevæger sig højt over hendes hoved. En samling af yngre mænd og kvinder, sandsynligvis nævninge eller advokater, der indtil nu er forblevet siddende, læner sig pludselig alle forover mod kvinden, idet hun kolliderer i armene på den yngre kvinde.

Konteksten er uklar, men senere i filmen finder vi ud af, at udraget, vi netop har set, stammer fra en krigsforbryderdomstol i Tokyo år 2000,<sup>62</sup> og at kvinden har været sexslave under den japanske hærs kontrol. Vi ser hele sekvensen med kvindens kollaps i ekstrem slow-

motion. Samtidig med at slowmotionbilledet sætter sig i os, hører vi en stemme sige: "Trying to figure out how I can talk about something that is unseen [...] that most people don't believe in [...] something really horrible." En anden stemme tilslutter sig den første: "I have been having very strange dreams. I had nightmares of being massacred." Den første stemme bryder frem igen: "The secret or the gap in one's speech gives rise to a ghost." Endnu en stemme introduceres: "En følelse som vrede og en følelse som sorg, og de følelser hænger sikkert også rigtig meget sammen, det er to følelser, som jeg har lagt mærke til." En fjerde stemme istemmer: "The part we didn't know that was hidden from us and forbidden from us."

Vi lytter til ordenes meningsindhold, men langsomt skifter ordene karakter, i takt med at stemmerne akkumulerer. Det bliver svært at skelne den ene stemme fra den anden. Forskellige sprog, dansk, engelsk, koreansk, kinesisk og accenter krydser over hinanden, fortætter lydbilledet, så kun fragmenter af sætninger kan høres, og ordene opløses til klangflader og rytmer, der bevæger sig frem og tilbage mellem hinanden. Brudstykker af ord dukker frem for at forsvinde tilbage i det flerstemmige kor af intensiverede stemmer.

### Stemmernes påtrængende nødvendighed

Når vi ser åbningssekvensen, den ældre kvinde med hånden bankende på hjertet og hendes krops efterfølgende kollaps, ser vi, hvordan smertelige minder af fortiede traumer manifesterer sig i kroppen. Hendes bevægelser demonstrerer mangelfuldheden i sproget. Idet hun skal afgive sit vidnesbyrd, kommer hendes ord til kort, og kroppen tager over. Bevægelserne udtrykker oplevelser og følelser, der har været fortiet og undertrykt i mange år, noget, der ikke kan beskrives fyldestgørende med ord. Nancy Rose Hunt, specialist i afrikansk historie og antropologi, skriver om akustiske registreringer som vidnesbyrd i forbindelse med massevoldtægter af kvinder i Den Demokratiske Republik Congo fra 1903 og frem til 2003 (Hunt 2008: 236). Hun advokerer for, at vi ikke blot lytter efter ordene, når vi lytter til vidnesbyrd, men at vi også lytter efter, *hvordan* ordene som materialitet indeholder lyde af skam og sorg. Således kan vi begynde at forestille os det, som det ikke er muligt at tale om (Hunt 2008: 242). I *The Woman, The Orphan, and The*



Da kvinden kolliderer i retssalen, hører vi et kakofonisk kor af stemmer i værket *The Woman, The Orphan, and The Tiger* (2010) af Jane Jin Kaisen og Guston Soudin-Kung.

*Tiger* oplever vi korets stemmer som en lydlig pendant til kvindens faldende krop. Koret bliver et poetisk vidnesbyrd, en måde at give stemme til det, der har været fortiet. ”I *The Woman, The Orphan, and The Tiger* er der en vægt på stemmerne som materialitet, tonalitet. Stemmerne tøven og påtrængende nødvendighed i at blive hørt,” fortæller kunstneren Jane Jin Kaisen (Jin Kaisen 2016). Vi lytter til koret. Stemmerne affektive klange og rytmer, der fremtræder og blander sig med hinanden. Den kakofoniske cementering af stemmerne i koret fører os mod en alternativ lytning af de ord, der bliver sagt. Den semantiske lytning træder i baggrunden til fordel for en affektiv lytning efter stemmerne tonalitet og materielle intensitet. Der er en indtrængenhed i stemmerne, som er ladet med betydning og relevans.

Korets mangfoldige stemmer skal ikke ses som et billede på kvindernes individuelle personlige historier. Koret peger ud over dem, rækker ud over det enkelte menneskes vidnesbyrd og breder sig ud som en kollektivitet. ”Det er meget på grund af det indhold, som *The Woman, The Orphan, and The Tiger* har, at vi valgte at starte filmen med et kor af stemmer, så der lige fra begyndelsen er en vægt på det kollektive frem for en individuel stemme,” siger Jin Kaisen og fortsætter: ”Det, vi ønskede med filmen, var at intervenere i tendensen til at læse disse fænomener som isolerede og i stedet se på de underliggende strukturelle forhold; den biopolitiske mobilisering af kvinder, der opstod i en nek-sus af kolonialisme, militarisme og patriarkat, ikke blot som et korean-sk problem, men som et transnationalt problem.” (Jin Kaisen 2016)

Koret manifesterer sig som et kollektiv af forskellige stemmer og materialiserer derved det strategiske, politiske slægtskab mellem be-folkningsgrupper i Sydkorea, som har været marginaliserede, og hvis historier har været fortiet. Ved at markere slægtskabet, et slægtskab, som består i ikke at blive talt om og ikke at blive hørt – et slægtskab af spøgelser i Sydkorea, som også vækker genklang i diasporaen uden for landets grænser – peger stemmerne ud over sig selv i en anfægtelse af en national diskurs af fortællelse og distance. Koret materialiserer igen-nem vokale praksisser ønsket om og kampen for at indskrive sig i den sydkoreanske nationale historie og selvforståelse.

I *The Woman, The Orphan, and The Tiger* skaber koret et dobbelt vidnesbyrd. Koret henvender sig til os med vibrerende intensitet.

Beder os lytte til stemmerne materialitet. Knyttede hænder. Et poetisk vidnesbyrd. Vi bliver, som lyttere, vidne til stemmerne kredsen om det, der ikke kan siges. De oplevelser, vilkår og følelser, som er forankret i kroppen; fortællelserne, undertrykkelsen og fremmedgørelsen. Det man ikke kan indfange med ord, som man må cirkle omkring, men som gennemsyrrer hele ens tilstedeværelse, *The Tiger*.

### Dialogen mellem værk og lytter

I *Emigrant, Miss no. 1, 2, 3* og *The Woman, the Orphan, and the Tiger*, tre danske værker med internationale dimensioner, er vi inviteret med i de refleksionsprocesser, der udfoldes omkring de følelsesmæssige, intime, kropslige, repræsentative og politiske dimensioner af stemmen. I dialogen mellem værk og lytter bliver stemmerne som metafor for identitet, repræsentation og kollektivitet undersøgt, konstitueret, anfægtet og nuanceret igennem poetiske iscenesættelser af vokale praksisser og udtryk. Vi inddrages i værkerne som aktive lyttere, medvirkende i værkernes meningsproduktion. Værkerne giver ikke nogen entydige svar, men stiller i stedet spørgsmål til lytteren: Hvilken betydning har stemmen som repræsentation og identitetsmarkør? Hvad sker der, når stemmen ikke længere udsiges fra en krop, men er medieret igennem teknologier eller systemer? Og hvordan kan man give stemme til alt det, der ikke kan tales om?

Stemmen som kunstnerisk medie låner sig ud til en form for dialogisk meningsproduktion i værket. Weidman skriver, at stemmen er et sted, hvor værdier og forståelser, følelser og æstetiske udtryk manifesteres, debatteres og anfægtes igennem kropslige vokale praksisser (Weidman 2014: 38). Vi skal lytte til værkernes mobiliseringer af stemmer som poetiske anfægtelser, undersøgelser og overskridelser af identitetspolitiske problematikker. Som lyttere er vi en væsentlig del af dialogen; det er igennem vores lytten til stemmerne i værket, at værket udfoldes. Derfor handler det ikke kun om stemmerne i værkerne, men også om, hvordan vi lytter til dem; dialogen mellem os og værkerne. Der ligger således en forpligtelse i relationen. Et ansvar for at lytte til stemmerne i værkerne; deres tegnværdi, deres politiske udsagn og deres vidnesbyrd.

### Stemmernes paradoks

I *VØICE* skriver Neumark, at stemmen tidligere er blevet opfattet som autentisk og nærværende, men at det nu, i kraft af nye teorier og værker, er muligt for os at høre stemmen på ny, som fundamentalt paradoksal, og ikke som noget, der frembringer en original tilstedeværelse, identitet og subjektivitet (Neumark 2010: xvi). Stemmens paradoksaltet består i, at stemmen er noget af det mest intime, der kan definere en person, samtidig med at den aldrig udelukkende definerer personen, men altid er ude i verden, som en interrelationalitet, der opstår mellem stemmen og lytteren (LaBelle 2006: 106). Stemmen overskrider de kropslige grænser, hvorved den klassiske subjektivitet defineres, og ligger i mellemrummet. Dette mellemrum er det, som *Emigrant*, *Miss no. 1, 2, 3* og *The Woman, the Orphan, and the Tiger* indkredser. De dialoger, som udfoldes mellem værkerne og os som lyttere, muliggør en oplevelse af stemmerne, vi hører, som fundamentalt paradoksale, flertydige og komplekse. Ved at diskutere stemmens materielle vokale praksis, og hvordan denne praksis relaterer sig til stemmen som metafor i værkerne, er det blevet klart, hvordan kunstnerne arbejder med stemmens paradoksaltet og kompleksitet som et medie, der kan italesætte identitetspolitiske spørgsmål, immigration, køn og undertryktes stemmer, uden at det kommer til at omhandle personlige historier eller specifikke menneskers oplevelser. I stedet, i kraft af stemmernes flertydighed, fremtræder stemmerne som en anfægtelse og kritik af vores oplevelse af stemmen som autentisk, tilhørende en identitet og subjektivitet. Værkerne destabiliserer identitetskategorier og sætter vores oplevelse af tilhørsforhold i et andet perspektiv, som samfundsproblematikker, der eksisterer på et globalt plan. Værkerne er internationalt orienterede og taler ind i den internationale samtidskunst, selv når de, som i *Emigrant*, handler om specifikt danske forhold. Dermed får de, i dialog med lytteren, koblet spørgsmål om identitetskategorier og tilhørsforhold med problematikker, der eksisterer på tværs af kontinenter, sprog og befolkningsgrupper.

Spivak, G.C. (1988) "Can the Subaltern Speak?" i: C. Nelson og L. Grossberg (red.) *Marxism and Interpretation of Culture*. Urbana: University of Illinois Press.

Wissendorff, A. (2011) "Bykunstner overraskede på Horsens Kunstmuseum", *Horsens Folkeblad*, 28. november.

## Samtidskunst og agens

### Litteratur:

Gell, A. (1998) *Art and Agency: An Anthropological Theory*. Oxford: Clarendon.

Jalving, C. (2011) *Værk som handling: Performativitet, kunst og metode*. København: Museum Tusulanums Forlag.

Latour, B. (2007) *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.

Nochlin, L. (1988) "Why Have There Been No Great Women Artists?", *Women, Art and Power & Other Essays*. New York: Harper & Row.

Schlosser, M. (2015) "Agency", i: E.N. Zalta (red.) *The Stanford Encyclopedia of Philosophy*. Tilgængelig på: <https://plato.stanford.edu/archives/fall2015/entries/agency/> [23. september 2018]

Soussloff, C.M. (1997) *The Absolute Artist: Historiography of a Concept*. Minneapolis: University of Minnesota Press.

## Klasseblindhedens diskurs. Om inklusion og eksklusion i socialt engagerede kunstneriske praksisser

### Litteratur:

Bishop, C. (2012) *Artificial Hells: Participatory Art and The Politics of Spectatorship*. London: Verso.

## Stemmernes politik i samtidskunsten

### Noter:

59 Kurator og kunstteoretiker, Irit Rogoff skriver i artiklen "We – Collectivities, Mutualities, Participations" fra 2004, at de betydninger, som et kunstværk producerer, er noget, der skabes og deles kollektivt gennem cirkulationer af mening og relationer, der udfoldes rundt om kunstværket. Denne kollektivitet konstituerer et 'vi' i relation til kunstværket, hvilket er det, der henvises til med det 'vi', som optræder i denne artikel.

60 Af bøger med stemmen som filosofisk og teoretisk omdrejningspunkt kan blandt andet nævnes Connor 2000; Cavarero 2005 og Dolar 2006.

61 På dansk kan navneordet oversættes til "frøken", verbet til "at ramme ved siden af", og af synonymer kan eksempelvis nævnes "kvinde, dame, donna".

62 Krigsforbryderdomstolen i Tokyo var ikke organiseret af den japanske regering, men af forskellige NGO'er og havde derfor ikke dømmende magt. Se Matsui 2001.

### Litteratur:

Butler, J. (1997) *Excitable Speech: A Politics of the Performative*. New York og London: Routledge.

Cavarero, A. (2005) *For More than One Voice: Toward a Philosophy of Vocal Expression*. Stanford: Stanford University Press.

Chion, M. (1999) [1984] *The Voice in Cinema*. New York: Columbia University Press.

Connor, Steven (2000) *Dumbstruck: A Cultural History of Ventriloquism*. Oxford: Oxford University Press.

Dolar, M. (2006) *A Voice and Nothing More*. Cambridge, MA: MIT Press.

Hamdan, L. A. (2014) "Aural Contract: Forensic Listening and the Reorganization of the Speaking Subject", *Cecura//Acceso*, efterår 2014, nr. 1, pp. 200 – 224.

Hunt, N.R. (2008) "An Acoustic Register, Tenacious Images, and Congolese Scenes of Rape and Repetition", *Cultural Anthropology*, vol. 23, nr. 2.

Jin Kaisen, J. (2016) Interviewet af Stina Hasse Jørgensen, 9. juni.

Kahn, D. (1999) *Noise, Water, Meat: A History of Voice, Sound, and Aurality in the Arts*. Cambridge, MA: MIT Press.

Kelly, M. og C. Widenheim (2011) *Dialogue*. Stockholm: Iaspis.

LaBelle, B. (2006) *Background Noise: Perspectives on Sound Art*. New York: Continuum Books.

Lysholt Hansen, N. (2016) Interviewet af Stina Hasse Jørgensen, 7. juni.

Matsui, Y. (2001) "Women's International War Crimes Tribunal on Japan's Military Sexual Slavery: Memory, Identity, and Society", *East Asia (Piscataway)* 19(4).

Neumark, N., R. Gibson og T. van Leeuwen (2010) *VØICE*. Cambridge, MA: MIT Press.

Pugholm, N. (2016) Interviewet af Stina Hasse Jørgensen, 4. juni.

Rogoff, I. (2004) "We – Collectivities, Mutualities, Participations", *Transart Institute*, 8. juni. Tilgængelig på: <http://www.transart.org/events-2016/files/2016/05/We-Collectivities.ROGOFF.pdf> [23. september 2018]

Weidman, A. (2014) "Anthropology and Voice", *Annual Review of Anthropology*, oktober 2014, vol. 43.

Young, M. (2015) *Singing the Body Electric: The Human Voice and Sound Technology*. Farnham, Surrey, Burlington: Ashgate.



# **The Bloomsbury Handbook of Sound Art**

**Edited by Sanne Krogh Groth  
and Holger Schulze**

BLOOMSBURY ACADEMIC  
NEW YORK • LONDON • OXFORD • NEW DELHI • SYDNEY

BLOOMSBURY ACADEMIC  
Bloomsbury Publishing Inc  
1385 Broadway, New York, NY 10018, USA  
50 Bedford Square, London, WC1B 3DP, UK

BLOOMSBURY, BLOOMSBURY ACADEMIC and the Diana logo are trademarks of  
Bloomsbury Publishing Plc

First published in the United States of America 2020

Copyright © Sanne Krogh Groth, Holger Schulze, and contributors, 2020

Cover image: *Klik* by Christian Skjødtt and Esben Bala Skouboe, Main Library and  
City Council Chamber of Aalborg, Denmark (2013). Photo courtesy of the artists

All rights reserved. No part of this publication may be reproduced or transmitted in any form or  
by any means, electronic or mechanical, including photocopying, recording, or any information  
storage or retrieval system, without prior permission in writing from the publishers.

Bloomsbury Publishing Inc does not have any control over, or responsibility for, any third-party  
websites referred to or in this book. All internet addresses given in this book were correct at the time  
of going to press. The author and publisher regret any inconvenience caused if addresses have changed  
or sites have ceased to exist, but can accept no responsibility for any such changes.

Whilst every effort has been made to locate copyright holders the publishers would be grateful  
to hear from any person(s) not here acknowledged.

A catalog record for this book is available from the Library of Congress.

ISBN: HB: 978-1-5013-3879-3  
ePDF: 978-1-5013-3881-6  
eBook: 978-1-5013-3880-9

Series: Bloomsbury Handbooks

Typeset by Deanta Global Publishing Services, Chennai, India  
Printed and bound in the United States of America

To find out more about our authors and books visit [www.bloomsbury.com](http://www.bloomsbury.com) and  
sign up for our newsletters.



# 4

## Vocalizing Dystopian and Utopian Impulses

### The End of Eating Everything

*Stina Marie Hasse Jørgensen*

---

## Introduction

How are dystopia and utopia vocalized in contemporary sound art? Dystopia and utopia have been important impulses for art to create future scenarios and imagining the world otherwise. Dystopia translates from Greek as “bad place” and is an undesirable imagined place. Characterized as a state of agony, dehumanization, and disaster, dystopia appears as an impulse in many artistic works. The word “utopia” translates from Greek as “not place.” Five hundred years ago, Thomas More proposed a conceptualization of utopia as a non-place where an ideal form of human organization could exist (More 2016). Richard Noble, professor in art history, writes in the introduction to the anthology *Utopias*, from 2009, about how the notion of utopia has been depicted and reconceptualized in art throughout time. He argues that one can find utopian impulses in many artworks in that they both “direct our attention to the realm of the political,” and are “oriented beyond existing conditions, sometimes to the future, sometimes to the past; it is art that asks us difficult questions about the conditions we live with and the potential we have to change them” (Noble 2009, 14). Noble’s main concern is how these impulses can be seen in visual art but leaves out the ways in which sound in art makes them heard.

In the following, I will discuss utopian and dystopian impulses in contemporary artworks and argue that they are closely connected, as if two sides of a single coin, since both can arguably be places that are imagined but not realized, pointing to political issues as something that extends from the present to the future or to the past. Here, I will explore how sounds, specifically voices, in an interplay with the visual dimensions of different artworks, make utopian and dystopian impulses heard. I will focus on the performativity of voices as artistic media in artworks. Sound studies scholar Norie Neumark writes in “Doing Things with Voices: Performativity and Voice” that when voices work performatively,

they don't just describe or represent something or someone, but perform and activate, for example feelings of intimacy or intensity. She states; "performativity suggests something that doesn't just describe or represent but performs or activates—acting as "a material force to change something." Performative voices can "*provoke, invoke, evoke, and convoke*" (Neumark 2010, 96) feelings in the audience as a performative action that enacts through the material qualities of the sounding voice, which goes beyond what the voices semantically say. Neumark states that "rather than call or call forth an audience as interactors, voice can instead performatively interpolate them as players and/or performers who bring the works into existence" (Neumark 2010, 97).

I will discuss how three contemporary artworks made by female artists from different parts of the world vocalize dystopian and utopian impulses by different means. My focus will be on how the performativity of voices as artistic media in relation to the visuality of the artworks makes dystopian and utopian impulses heard, and how this can engage the audience in the complexity of political issues and create performative experiences that direct the audience's attention to imagine the world otherwise. The works are all unfolded through time and include vocal performances, yet in different ways, through different media, and with attention to different political issues that concern voices of the otherwise unheard—through screams, dialogues, or meditative recitation.

First, I will discuss how the animated video *The End of Eating Everything* from 2013 by the Kenyan-born artist Wangechi Mutu performatively uses a monstrous scream as an emblematic knot of synchronization in the interplay with the visual elements in the video in order to engage the audience. Then, I will account for the ways in which the play with shadows and voices in a poetic dialogue includes the visitor in the video/shadow play *Transgressions* from 2001 by Indian artist Nalini Malani. Lastly, I will examine how the meditative recitation by the Danish artist Nanna Lysholt Hansen in the speech performance *Dear Daughter/Anatomy of the Chthulucene #6* from 2018 is performatively playing with dubbing, playback, and loose synchronization of voices in relation to the naked body in the performance space, engaging the audience through a discomfiting uncertainty about the connection between the naked body and the voices speaking. The three artworks presented map out different means by which the artworks make use of the performativity of voices in relation to the visual expressions involving the audience in the issues pointed at in the artworks.

## ***The End of Eating Everything***

The 8-minute video animation *The End of Eating Everything* was made by internationally acclaimed contemporary artist Wangechi Mutu (born 1972). Mutu is known for her paintings, sculpture, and videos exploring different aspects of the female body as subject in terms of gender constructs, cultural trauma, and environmental disaster. She has exhibited at museums worldwide, including Tate Modern in London, Centre Georges Pompidou in Paris, the Whitney Museum of American Art in New York, and the Museum of Contemporary Art in Los Angeles among others.

*The End of Eating Everything* was commissioned as part of the exhibition *Wangechi Mutu: A Fantastic Journey* at Nasher Museum of Art at Duke University, North Carolina in the United States. In the video animation, screaming birds fly in formation across the sky of a gray dystopian landscape. A close-up of a black woman's face (the musician Santigold) depicts her looking at the birds quietly, with glowing stones or dandruff in green, pink, and purple at her temples. As the camera zooms out it becomes visible that the woman has Medusa-like dreadlocks swirling around her face as if they were made of eel-like creatures with a life of their own. It is revealed, as the camera zooms even further out, that the Medusa-haired head is not attached to a human body, but to a floating earth-body. Arms, legs, and industrial wheels are sticking out of the floating earth-body, making the hybrid creature look like one of the figures found in Hieronymus Bosch's visions of hell from the triptych, *The Garden of Earthly Delights* (1490–1510), placed in an enchanted world that is both ancient and futuristic at the same time. Below the disturbing imagery of the hybrid Medusa creature a fragment of the textual narrative unfolding throughout the animated video reads "Hungry, alone and together" (Mutu 2013). The birds scream and the hybrid Medusa creature sniffs in the air as if smelling a meal in front of it. They are all hungry, alone, and together. A dissonant drone, which sounds like a time-stretched scream, hovers in the background, underlining the dystopian atmosphere. The hybrid Medusa creature pours out a powerful and horrific monstrous scream before she attacks the flock of screaming birds and eats them. In the air/water traces of blood from the birds are floating. Then, polluted fog rises from the hybrid Medusa creature, which causes it to implode into smaller pink-purple cell-like replicas of a hairless Medusa head, all smiling ecstatically.

## The Scream of the Hybrid Medusa

In many ways, it seems as though Mutu's hybrid creature embodies the Medusa from Greek mythology, a chthonic monster full of female rage from an archaic world, with living snakes in place of her hair. In Mutu's work, however, the hybrid Medusa is not just operating in the archaic world, but is just as much a monster living out its female rage in a time that merges past, present, and future. The roaring scream of the hybrid Medusa captures the dystopian impulse materially in a horrifying moment. The monstrous scream also convokes, through the material qualities of the performative voice, the audience's experience of the merging of temporalities. The visual animation of the dramaturgy in the video supports this experience of the scream as performative; the narrative converges in the moment of the roaring scream and radiates from it. In this way, the scream is what the film theorist and composer Michel Chion has coined as a punch sound, which is "the audiovisual point toward which everything converges and out of which all radiates" (Chion 1994, 61). The punch sound is an emblematic knot of synchronization where the visual imagery and the sound converge, becoming at one and the same time separate. The scream is a punch sound in that it is, at one and the same time, converging and separating from the visual depiction of the hybrid Medusa. The horrific scream converges with the imagery

of the composite Medusa combining human and nonhuman elements. Yet, at the same time, it is separated from depiction of the hybrid Medusa, as it sounds like it comes solely from a human body, which creates a chilling ambivalence of the monstrous creature as an uncertain other—both human and nonhuman. It is in this conversion and separation between the voice and the imagery, which creates an unplaceable monstrous otherness, that there is a generative potential.

## Listening to the Monstrous

Philosopher and contemporary art theorist, Bojana Kunst, writes in “Restaging the Monstrous” about how the monstrous has a generative potential. She argues that the monstrous is other than human, but also a symptom of rethinking the human. Kunst discusses the political preoccupation with the monstrous and finds that it is “in the ways in which the other (animal, slave, machine, woman, etc.) is continuously humanized to reflect back the face of ‘our’ own (white, western and male) humanity” (Kunst 2008, 215). Kunst further writes:

the monstrous becomes the “ever present possibility to destroy the natural order of authority” not because it is some externalized other which has to be swept into the arms of regulating order, but because it is the constant production of otherness in the very *human* being, so that the human can recognize and define itself. (Kunst 2008, 215)

In *Postopera: Reinventing the Voice-Body*, the musicologist Jelena Novak links the generative potential of the monstrous as an otherness, as it is conceptualized by Kunst, to that of the monstrous voice, which has a “terrifying friction of inhuman noise” (Novak 2015, 67). Following Novak, the punch sound, which here is the performativity of the voice in relation to the visibility of the hybrid Medusa’s body, creates a monstrous



**Figure 4.1** Nalini Malani, *Transgressions*, 2001. Courtesy Stedelijk Museum, Amsterdam, Netherlands.

otherness, which has a generative potential. Said another way, the hybrid Medusa in the visual imagery arguably produces an otherness, but through the affective intensity of the monstrous voice, time-stretched to linger as a friction between inhuman noise and the performative human scream, this otherness is not external to the human, but instead creates an experience of the constant production of otherness in the human. This internalization of an otherness in the human enables the audience to recognize themselves in connection to the monstrous. Kunst writes that the monstrous is “materializing the very potential of hybridity to open up possibilities for a different world” (Kunst 2008, 220). The monstrous scream interpolates the audience and engages them in the artwork through their listening to the hybrid monstrous Medusa (as part human and part Earth) as an otherness that can destroy the existing the order of things, pointing at the need for a new moral compass that can change the behaviors and actions of humans (pointing at the audience). The performative punch sound is a material force, and a catalyst for the audience’s reflection: Can the Earth become anything else than a monstrous Medusa full of rage in the age of eating everything? Is it possible to share and care for others, nonhuman as well as human, in the times of the Anthropocene, the period during which human activity has been impacting Earth’s geology, climate, and environment?

There is a duality here connecting the dystopian and the utopian impulse in the audience’s audiovisual experience of the video work; condensing the critique of the time of the Anthropocene and the consumption and exploitation of the Earth in the past, present, and potentially the future, as well as creating a space for the audience to reflect and imagine otherwise, other ways of consuming, eating, and caring for nonhumans as much as humans.

## Transgressions

The video/shadow play *Transgressions* is made by the renowned contemporary Indian artist Nalini Malani (born 1946), who is considered a politically motivated artist bringing attention to issues of transnational politics and issues of gender and postcolonial history, using different media such as painting, theater, and installation projects. Malani has exhibited all over the world at places such as Kiran Nader Museum of Art in New Delhi, the New Museum of Contemporary Art in New York, and the Institute of Contemporary Art Indian Ocean in Mauritius.

*Transgressions* directs the audience’s attention to the complexities of political issues through the use of performative voices and their interplay with the visual elements of the work, in a way that stretches—or transgresses as the title of the artwork suggests—beyond existing conditions, to the past and to the future, incorporating a duality of dystopian and utopian impulses, in a way that is similar to Mutu’s video animation. *Transgressions* is like *The End of Eating Everything*, a work that unfolds over time, however the movements of the images are here animated mechanically by rotating motors onto video projection, and the voices operate both on a semantic level and a material, performative level. Although Malani’s work is concerned with the complex issues of consumption, just as in *The End of*

*Eating Everything*, *Transgressions* connects it to the problems of the global economy and the history of imperialism.

In the following, I will point to the ways in which *Transgressions* engages audiences in these complex political issues through poetic conversations, which unfold as affective dialogical fragments entangled with the visual imagery in the video/shadow play. Neumark argues that the way voices performatively engage audiences is by evoking feelings in the audience through the material qualities of the sounding voice, for example through an aesthetics of intimacy and intensity. The voices, Neumark argues, can call the other into an intimate relationship. “This happens not by speaking of intimacy, but through vocal qualities and vocal performance—through the performativity of the voice” (Neumark 2010, 96). In my discussion of *Transgressions*, I will describe how the performativity of voices can be said to interpolate audiences through an aesthetics of intimacy and intensity. This performativity of voices lets otherwise unheard voices of women and children poetically embody current political issues of globalization and consumption. This enables audiences to reflect upon the problems of today and the future—mobilized through references to colonialism, ancient mythic characters, mobile phone companies, and linguistics.

## “I Speak Orange”

*Transgressions* is a combination of different media in a distinct hybrid genre, the video/shadow play. The installation is created as an immersive environment with four transparent, slowly rotating Mylar plastic cylinders. On the cylinders there are various scenes that combine Indian and Greek mythology with historical events, especially the English colonization of India, in a patchwork merging history and fiction, such as the depiction of the goddess Kali holding the decapitated heads of English colonizers in her many arms. This, as well as illustrations of different organs from the human body and animals, such as monkeys and lions, is painted in the traditional Indian Kalighat style. The scenes and figures are painted from the inside of the cylinders with the so-called “reverse painting” technique. Three video projections are traversing the cylinders showing the skin of a Caucasian male as a reference to the postcolonial history of India. When the reversed paintings from the cylinders cast shadows on the walls in the gallery space it is as if they drift across the Caucasian male’s skin, like animated tattoos that dynamically change size and merge with other shadow figures. The shadows from the paintings on the rotating cylinders are framing and enclosing the visitors in the installation. In fact, the visitors become part of the shadows in the installation having their own figures casting shadows on the walls, merging with the figures from the video/shadow play. In this way, visitors are affectively engaged and implicated in the video/shadow play.

Walking inside the video/shadow play visitors can hear the ambient soundscape of a sitar playing the same chords on a loop as a meditative backdrop to voices speaking in

poetic language, which is seductive and dream-like. A female-sounding voice states a once-utopian aspiration: “We had everything, everything, everything before us” (Malani 2001). Yet the belief in better times to come is punctured by the way the statement is spoken in the past tense, which makes the visitor aware that this aspiration is not applied to the present moment in time, serving as the future (where everything was possible) to the moment the statement was made. In this way, the entanglement of present and future times is introduced as something that is inevitably transgressing the boundaries of separated temporalities. This transgression of linear time is also stressed by the cyclic time manifested in the rotating cylinders and the looped sound of the sitar on the soundtrack. It becomes clear that, for the speaking voice, time also equals money when it starts to list prices on mobile phone times: “Air time Rupees one forty nine” (Malani 2001). Then the following lines are rhythmically spoken: “I speak Orange, I speak blue, I speak your speak just like you” (Malani 2001). The word Orange refers to one of the mobile phone providers in India at the time the video/shadow play was created. The reference to the brand Orange makes a gesture toward the processes of globalization and consumerism that continuously transform the conditions of people living in India and elsewhere. Time, not only in relation to the time of labor, but also to the time of consumption, such as speaking on the phone, is connected to a value and enrolled in the monetary economy.

The work makes evident that not only time is given a value but so too are languages, and it illustrates how some languages are seen as worth more than others. This is emphasized when a technologically modified voice, which sounds like that of a little girl, says: “Amma, please send me to English school / and Amma, she really was no fool” (Malani 2001). The child’s plea to her mother proposes that the English language is required in order to succeed and get a job in the society of the future. In India, English is also to a large extent regarded as the language of global consumerism, which is emphasized when a female-sounding voice says: “So la ti do / I do, I do in high heel shoe / I do, I do want to be like you / I orange, I blue / I do, I do” (Malani 2001). Languages have been given a value and, in the same way as time, are enrolled into the monetary economy. Meanwhile the voices speak in English on the installation’s soundtrack, the video projections display Indian scripts in local Indian languages: Urdu, Telugu, Tamil, Malayalam, Gurumukhi, Bengali, Gujarati, and Marathi. The Indian languages slowly fade into the ground as if symbolizing that they will soon disappear as a consequence of globalization, wherein the dominant spoken language is English—the valued language of multinational companies and the future of consumerism. Being and speaking English “just like you” is the means of access to money and therefore also the future. As art historian Mieke Bal states: “the thought that English opens all doors and that without English, one is doomed to poverty, has bound language to money, and money to English. This is the cultural loss of globalization” (Bal 2016, 253). This dystopian cultural loss, which is stressed by the video projections, concerns the increasing standardization of languages and expressions in the quest for globalization, on behalf of the complexities and troubles of speaking and translating many different languages.



## Giving Voice to the Shadows

The vocal exchange in *Transgressions* can be experienced as a poetical portrayal of a mother and her child, talking about what appear to be familiar things, such as choice of school and shopping. The voices sound as if they have been recorded close to the microphone. Here Malani is working with the potentials for intimacy without implying some singular identity but what can instead be experienced as enculturated bodies. Neumark explains the enculturated body: “Emerging from the body, voice is marked by that enculturated body. That is, *embodied* voices are always already mediated by culture: they are inherently modified by sex, gender, ethnicity, race, history, and so on” (Neumark 2010, 97). The “I” in “I speak orange” or “I do want to be like you” creates an intimacy as a vocal presence, modified by history, gender, and ethnicity. Words such as “Amma,” which can be translated to the English word “mom,” and the references to “Rupees” and “Orange” situates the voices as coming from enculturated bodies speaking from within the site of India where the postcolonial and capitalist struggles are unfolding. The “you” from the sentence “I do want to be like you” performatively interpolates the visitors as performers, inscribing them into the work by calling on the visitor as an enculturated body, as a representative of the ones mastering the English language, and as somebody who is part of the consumer culture.

The sound of the voices in *Transgressions* calls our attention to the materiality of the voices, for example the rhythm and tone in which the words are pronounced, and not just the words they say. Especially the childish voice questions the notion of identity through the materiality of the voice; it is hard to place, seeming as a displaced floating voice with no body. It is a technologically modified, intense, high-frequency-sounding voice, ambiguous in its appearance to the extent that it poses questions, or as to whether it is the voice of a child at all. The voice is not simply clear-cut child or not-child, it sounds in-between—as an Other that is hard to locate as a singular identity, but still an enculturated body. In this way Malani states that *Transgressions* is not about identity politics, but rather about structural problems connected to the ways in which the systems of globalization and consumerism enables or disables voices of Others outside the dominant currencies and languages of these systems to be heard.

It is here the utopian impulse of the work is unlocked, creating a situation or a space where the otherwise unheard Others, the ones in the shadow, are listened to, pointing at the ones potentially excluded from the systems of globalization and consumerism, drawing lines to the times of colonialism as another system constituting the voices of the colonial subject as Other. This connects to postcolonial theorist Gayatri Chakravorty Spivak’s category of the “subalterns”—those whose utterances cannot be heard by the system, owing to their social positions and the ways the system enforces its hierarchy. Subalterns are present but remain unheard in a global context. The Other is always in the shadow (Spivak 1994). In *Transgressions*, the Other is the child not able to speak the regional languages of where they were born, they speak English instead, because it is seen as the language of the future and the gateway to becoming successful in the



globalized world. In this way, the performative voices that speak in the installation give voice to the shadows and express a critique of past colonialism just as much as do the linguistic expansionism and imperialism of English as the language of wealth in the globalized world of the present and the future.

Neumark writes, the voice can act as a material force, a catalyst that can open the audience to experience and reflect upon political issues in a way where distant viewing and listening is impossible. The voices and images in the video/shadow play complement each other as assemblages, fragmented narratives, and storylines. Memories, imagination, dreams, and different temporalities are not clearly distinct in Malani's work, but come together through the interplay of the voices, video projections, paintings from the cylinders and their shadows, and the bodies of the visitors. They are all oscillating between appearing and disappearing, emerging and vanishing, manifesting an affective ambiguity as material yet ephemeral. The intimacy created by the vocal interpolation of the visitor, as well as the incorporation of the visitors' shadows, makes *Transgressions* a very physical and affective experience of the otherwise abstract politics of globalization and the world of consumerism. Malani is placing the visitor in a double role: as "you," the one who speaks English and does not understand the regional Indian languages, and as the role of the Other, the shadow on the white skin of the Caucasian male merging with the shadows from the cylinders. *Transgressions* both displays a dystopian impulse and a utopian impulse. The work is creating, on the one hand, a dystopian scenario where only English, as a profitable language in the global economy and rise of global consumerism,



**Figure 4.2** Nanna Lysholt Hansen, *Dear Daughter/Anatomy of the Chthulucene #6*, 2018. Courtesy of Nanna Lysholt Hansen.

is spoken, and less profitable languages, such as Indian languages, are forgotten and not spoken any longer. At the same time, on the other hand, the installation creates a utopian alternative space—giving voice to the otherwise unheard enculturated Others. *Transgressions* engages the visitors to listen to these voices and affectively implicates them in the situation unfolding in the video/shadow play, by placing the visitors in this ambiguous double role of Other and self.

## ***Dear Daughter/Anatomy of the Chthulucene***

The speech performance *Dear Daughter/Anatomy of the Chthulucene* #6 was created by contemporary Danish artist Nanna Lysholt Hansen (born 1980). Lysholt Hansen's artistic practice is concerned with the interplay between the female body as subject and object in connection to technologies. Lysholt Hansen has exhibited at places such as I: Project in Beijing, The Saint-Valentin Espace d'Art in Lausanne, The Pratt Manhattan Gallery in New York, and Kunsthal Charlottenborg in Copenhagen. Just like *The End of Eating Everything* and *Transgressions*, Lysholt Hansen's speech performance displays a dualism where a utopian and dystopian scenario are exhibited at the same time. All three works point to dystopian possible futures or pasts of disaster, the extinction or suppression of voices of Others, often women, in global consumerism and production. Yet, at the same time, the works all display a reimagination of another world where the voices of women and Others, the otherwise unheard, are heard.

*Dear Daughter/Anatomy of the Chthulucene* #6 was performed as part of the exhibition *Shaking the Habitual* concerning feminist utopias at the gallery space named *meter* in 2018. The performance is the last part in Lysholt Hansen's project *Dear Daughter*, which consists of a series of three performances. All three speech performances are recitations of letters spoken aloud to her daughter. In all the three of the *Dear Daughter* performances Lysholt Hansen interweaves, in her spoken letter to her daughter, her own experiences and thoughts on crucial elements in female life, such as pregnancy, birth, mothering, and female sexuality, with text fragments from feminist literary sources. The first series of performances, *Dear Daughter/Organic Cyborg Stories (After Donna Haraway)*, were first performed in 2013 as a letter to Lysholt Hansen's unborn daughter. The second part of the performance series, first performed in 2014, *Dear Daughter/Motherboard Theories of Evolution (w/ Braidotti, Plant et aliae)*, was a letter to Lysholt Hansen's newborn daughter. *Dear Daughter/Anatomy of the Chthulucene* (O' Connel Oh Oh Haraway) was first performed in 2016 as a letter to Lysholt Hansen's almost-speaking daughter and is about sex and sexuality. Here, Lysholt Hansen weaves the biologist and feminist thinker Donna Haraway's concept of the "Chthulucene," together with anatomic descriptions of the clitoris by female urologist Helen O'Connell as well as Lysholt Hansen's own experiences with mothering and female sexuality.

At *meter*, a giant clitoris sculpture made from white Styrofoam, from Lysholt Hansen's sculpture series *The Chthonic Ones/Anatomy of the Chthulucene* (2016–2018), stretched

its tentacular structure out in the gallery space. Four palettes made of something that looked like soft clay with neon colors on them were placed on the floor around the clitoris sculpture. Approximately twenty people were seated on a ramp or standing up against the walls of the small gallery space. Lysholt Hansen sat in a mermaid posture, completely naked except for a headset worn above a nylon sock-mask, which covered her head. She had a small speaker reciting the letter to the daughter that she moved around with her as she made her way down the ramp toward the clitoris sculpture. The ramp was not very broad, so Lysholt Hansen was more or less in between the legs of the audiences on the ramp. Lysholt Hansen painted child-like drawings on her naked body with loose lines of the luminescent neon colors from the palettes on the floor. It was a very intimate bodily experience with Lysholt Hansen's body almost touching the audience as she moved. At the same time, the performance seemed cold and distanced, which was stressed by the way that Lysholt Hansen had covered her head with the nylon sock-mask, the headset, and the speaking voice from the speaker that seemed to instrumentalize and alienate the naked body. However, the voice heard seemed intimate and calm, speaking slowly as if delivering a meditative guidance. Deep breaths and intimate sounds from the mouth are sometimes heard on top of the soundtrack bringing associations to new age music with long drones and harmonic tones slowly unfolding. The calm intimacy of the naked female body, the meditative voice, and the instrumentalization and alienation of the same body through the mediation of the voice through the speaker and the disconcertingly covered head are reflected in the semantic content of the letter.

Lysholt Hansen uses recent anatomical descriptions of the clitoris, a part of the female body that was mapped in detail for the first time by O'Connell in 2005. Throughout the performance it became clear that this mapping of the clitoris is part of an activism that aims to disseminate knowledge about the clitoris, and the sexuality that comes with it, because for long periods of time it has been suppressed and not spoken about. In the performance we are also told that the majority of the clitoris's anatomy is chthonic, or underground, stretching tentacularly into the body, linking Lysholt Hansen's description of the anatomy of the clitoris to Haraway's conceptualization of the Chthulucene in *Staying with the Trouble: Making Kin in the Chthulucene* (2016). This link was also made visible in white Styrofoam sculpture at the center of the performance. In the performance, the chthonic structure of the clitoris was also connected to the narrative frame of the Chthulucene, which is a conceptualization of our current time as a place where past, present, and the yet to come are entangled. Haraway describes the Chthulucene as the time of the chthonic ones, which are "beings of the earth," to which the organ of the clitoris can be said to belong, which are "both ancient and up-to-the minute" (Haraway 2016, 2). This notion of entangled time is crucial in Haraway's investigation of the possibilities of moving on from "disturbing times, mixed-up times and turbid times" (Haraway 2016, 1). As an alternative to the current conceptualization of our time as the Anthropocene where an onrushing destruction is taking place, manifesting a kind of dystopia, Haraway wants to bring back attention to the possibilities of feminist partial healing and making kin, and proposes a conceptualization of our current epoch as the

Chthulucene: a period where the sense of ongoing living with each other is unfolded in places where times and matter are entangled (Haraway 2016). Haraway's conceptualization of the Chthulucene—the entanglement of times and matter, where feminist partial healing and making kin is possible—is also embodied in Lysholt Hansen's speech performance through an interplay between technology, body, and performative voices that display a polytemporality and an aesthetics of intimacy, intensity, distance, and alienation creating both dystopian and utopian impulses.

## Zones of Entanglement between Machines and Organic Selves

The first sentences in Lysholt Hansen's *Dear Daughter/Anatomy of the Chthulucene* condenses how she combines her own experiences of motherhood in relation to her daughter with the anatomic descriptions of the clitoris and Haraway's notion of the entanglement of times in the Chthulucene: "Now we play / We are glitter, pink, purple, coral and slime / This which I am going to tell you reaches into your biology, your genealogy and from there stretches tentacularly and technologically into the Chthulucene, the future we are becoming. Again. Now" (Lysholt Hansen 2018). The linguistic meaning of the letter that Lysholt Hansen performs matters, but so too do the ways in which she materially performs it. Although Lysholt Hansen is corporeally present, her naked body painted with neon colors live in the gallery, her presence there is also mediated and estranged. The voice heard reciting the letter is not mainly coming from Lysholt Hansen's mouth as she sits on the gallery floor, but from the speaker she is moving with her whenever she moves her body. The dissociated voice heard in connection to Lysholt Hansen's slowly moving body destabilizes the experience of the "here and now" of the performance and creates questions as to whether the performance is personally authored, leaving the audience with a titillating discomfort.

The use of the dissociated voice from the speaker moving around the gallery space together with Lysholt Hansen's audiovisual presence makes sensual the zones of entanglement between machines and organic selves. In the speech performance Lysholt Hansen switches between playback, dubbing, and loose synchronization in order to attach the speaking voice from the loudspeaker to the performing body in the gallery, leaving the audience with an unsure feeling of from when and from whom the voice emanates. It is an ambiguous body-voice-technology relation unfolding in the performance. In the following, I will discuss why the alienating and destabilizing interplay between voice, technology, and body is crucial in making Lysholt Hansen's speech performance become a place where both dystopian and utopian impulses are exhibited, generating a space where conventions and norms are investigated and different social structures and interrelations are vocalized, enabling the audience to imagine otherwise.

## Dubbing, Playback, and Loose Synchronization

The tradition of dubbing, synchronization, and playback is one frequently used in film and it has been written about by Michel Chion. He has discussed the differences between body-voice relations with regard to these.

Chion writes that in playback performances “there is someone before us whose entire effort is to attach his face and body to the voice we hear. [...] Playback marshals the image in the effort to embody” (Chion 1999, 156). The playback technique is used when there is a recorded voice and miming body on stage. It is when Lysholt Hansen moves her lips behind the nylon mask as an effort to attach her face and body to the voice we hear. Chion points to the synchronization process that happens in the playback of the voice with precise movements of the lips, in order to simulate that the voice heard comes from the mouth moving on film. He writes that this is a representational procedure: “We take this temporal co-incidence of words and lips as sort of guarantee that we’re in the real world, where hearing a sound usually coincides with seeing its source” (Chion 1999, 128–9). Hearing the speaking voice from the speaker in Lysholt Hansen’s performance coincides with seeing the movements of the lips moving in real time behind the nylon mask on the neon-painted body. The voice coming from the speaker is in this way attached to the neon-painted body performing live in the gallery space. However, it is only sometimes that the articulation of the words heard from the speaker accords with the lip movements of the mouth behind the mask. Other times, the mouth behind the nylon mask is whispering the words in synchronization with the words uttered from the speaker, which appears to double the voice as if it both belongs and doesn’t belong to the performing neon-painted body. This is the effect of dubbing the vocal performance.

Chion writes that dubbing is when “someone is hiding in order to stick his voice onto a body that has already acted for the camera” (Chion 1999, 156). In Lysholt Hansen’s performance the dubbing is a way to stick the naked body’s voice onto a speaking body that has already acted for the voice recorder. It is a doubling of speaking bodies where one body is only heard and never seen. In film, Chion argues, dubbing can help characters to appear more realistic and comprehensible, creating the illusion of wholeness between image and sound. In Lysholt Hansen’s performance, however, the gap between the body and the technologically distributed voice destabilizes the experience of the performing naked body and the speaking body as a whole. Lysholt Hansen further emphasizes the destabilization by performing a loose synchronization, by a delayed performance of the words heard from the speaker, or by only moving the lips behind the nylon mask to some of the words heard. It stresses the fictitious world of the performance space. Michel Chion notices that “in general, loose synch gives a less naturalistic, more readily poetic effect” (Chion 1994, 65).

The use of playback, dubbing, and loose synchronization techniques feeds an experience of the body seen and the voice heard as both connected and disconnected at the same time.

It is a performance playing with the expectations of the audience, performing a complex and ambiguous relationship between body and voice, image and sound, live, recorded, and projected. When Lysholt Hansen uses the playback and dubbing techniques in her performance, it is not to make the temporal relationship between voice and body seem realistic to the audience, in the way Chion argues that the techniques are used in cinema. It is rather used to destabilize what the audience is experiencing, displacing the here and now of the voice-body relationship into a polytemporality.

## Vocal Kinship

The polytemporality of the *Chthulucene*, where past, present, and the yet to come are entangled, is manifested both in the semantic meaning of words told and in the vocal performances of the voices heard in *Dear Daughter/Anatomy of the Chthulucene* #6. Lysholt Hansen play with the synchronization of body-voice-technology through the use of playback, dubbing, and loose synchronization. This opens for a polytemporality that unfolds in the voice performance and materiality during the performance time. The merging of time is also stressed in the semantic dimension of the work, where voices of Haraway's writings in 2016, O'Connell's text from 2005, and Lysholt Hansen's script from 2017 are fused. The polytemporality and entanglement of times is also quite literally stated in the script to the speech performance: "sym-poetically we have all melted together. / Over time. / Over time / To different times" (Lysholt Hansen 2018). The speech performance does not take place solely in the present, but instead takes place in a polytemporality, where an entanglement of past, present, and the yet to come is merged.

The polytemporality in the performance creates an alternative space where the dichotomy between the utopian and dystopian impulses are dissolved throughout the temporality of the speech performance. The vocal entanglement of past, present, and future in the performance can be argued to unfold a dissociation of the voice-body relationship, creating a dystopian alienation of the female body and the notion of mothering. This is also stressed by the nylon sock-mask's objectification of the naked female body, with the neon-colored, child-like lines drawn onto it, visually manifesting how female sexuality, caring, and mothering have been, are, and possibly will be, suppressed and muted.

At the same time, the performativity of the voice and body enact and make evident a caring and mothering between multispecies, rather than just narrating it. In the performance, the caring and mothering is intimately unfolded between Lysholt Hansen and the voice from the speaker, the daughter, and the audience—between the organic and the technological, the human and the nonhuman. The performance presents the zones of entanglement of both time and interrelations. In this way the performance can be said to demonstrate the utopian impulse. The polytemporality of the entanglement of times emphasizes Lysholt Hansen's performance as an alternative place where utopian impulses can unfold—where voices of the otherwise unheard can be listened to, and where imagining otherwise is intimately acted out. Lysholt Hansen ends her speech



performance by opening up the relationship between mothers and daughters and, in this way, including the audience as daughters in a feminist partial healing, making kin and caring for the chthonic: “Dear Daughter, / Here is no mother. Here is mothering / here is caring for others, caring for the chthonic, for the self-with-others. Care. Dear Daughter. The future is female. / Now you take over” (Lysholt Hansen 2018). It is almost therapeutic listening to the otherwise silenced history of female sexuality, to the traumatizing and dehumanizing effects of the ways in which female sexual, tentacular being has been repressed, especially because Lysholt Hansen, through her performance, also creates an alternative space where the clitoris and female sexuality are not repressed, giving voice to the chthonic ones, a space where intimacy is fostered through multispecies’ caring and mothering.

## Conclusion

In this chapter, I have discussed how utopian impulses have been vocalized in different ways in the three artworks Wangechi Mutu’s animated video *The End of Eating Everything*, the video/shadow play *Transgressions* by Nalini Malani, and Nanna Lysholt Hansen’s speech performance *Dear Daughter/Anatomy of the Chthulucene #6*.

Each in their own way, the three artworks display a dualism where the utopian is inside the dystopian or the dystopian is inside the utopian, unfolding throughout the temporality of the artworks. In both Wangechi Mutu’s *The End of Eating Everything* and Nalini Malani *Transgressions*, dystopian scenarios are exhibited by audiovisual means as dehumanization and disaster through global consumerism and eating everything, or as destruction of culture and languages in the rise of English as the imperialistic language of global consumerism. At the same time, both *The End of Eating Everything* and *Transgressions* create utopian spaces where the voices of women and Others are heard; the scream from the hybrid Medusa and the voices of the enculturated bodies of Indian mothers and children speaking English instead of Indian languages. In these works these voices, otherwise repressed as chthonic monsters and Others, can scream and speak in a way that points to the complexities of political issues in a quest to prompt the reimagination of another world and their role in it. The speech performance *Dear Daughter/Anatomy of the Chthulucene #6* by Nanna Lysholt Hansen also holds the dualism of dystopian and utopian impulses, as the other artworks do as well. On the one hand, the work holds a dystopian scenario in its alienation and objectification of the female body and the notion of mothering. On the other hand, it creates a utopian space through giving voice to women, capturing the intricacies of political issues, and it creates an alternative space for experiencing an imaginative space where voices of the otherwise unheard are made audible.

The three artworks can be said to vocalize dystopian and utopian impulses in different ways using different media and vocal techniques. Through an entanglement of history, fantasy, memories, dreams, and fictional and real-time spaces, the audiovisual narratives and performances enable affective experiences of political issues that stretch out in relation

to the past and the future. In this way, the artworks present, in each their own way, alternative and imagined places in which real conditions, values, and positions are investigated. The artworks' visual elements, in their interplay with the performativity of voices as an artistic medium, exhibited as screams, poetic dialogue, or meditative speech, act as material forces or catalysts interpolating and opening the audience to experience concrete yet imagined dystopian and utopian places and times entangled in each other, and to reflect upon how things are, have been, and could be otherwise.



## Bibliography

Bal, Mieke. *In Media Res: Inside Nalini Malani's Shadow Plays*. Ostfildern: Hatje Cantz Verlag, 2016.

Chion, Michel. *Audio-Vision: Sound on Screen*, trans. Claudia Gorbman. New York: Columbia University Press, 1994.

Chion, Michel. *The Voice in Cinema*, translated by Claudia Gorbman. New York: Columbia University Press, 1999.

Haraway, Donna. *Staying with the Trouble: Making Kin in the Chthulucene*. Durham: Duke University Press, 2016.

Kunst, Bojana. "Restaging the Monstrous." In *Anatomy Live: Performance and the Operating Theatre*, edited by Maaike Bleeker, 211-222. Amsterdam: Amsterdam University Press, 2008.

Lysholt Hansen, Nanna. *Dear Daughter/Anatomy of The Chthulucene (O'Connell, Oh Oh Haraway)*. Berlin: Laboratory for Aesthetics and Ecology, 2017.

Lysholt Hansen, Nanna. *Dear Daughter/Anatomy of the Chthulucene #6*. Speech performance, 2018.

More, Thomas. *Utopia*. London; New York: Verso, 2016.

Neumark, Norie.. "Doing Things with voices: Performativity and Voice." In *Voice: Vocal Aesthetics in Digital Arts and Media*, edited by Norie Neumark, Ross Gibson and Theo van Leeuwen, 95-118. Cambridge, Massachusetts: MIT Press, 2010

Noble, Richard. "Introduction//The Utopian Impulse in Contemporary Art." In *Utopias*, edited by Richard Noble, 12-19. London: Whitechapel Gallery; Cambridge, Massachusetts: MIT Press, 2009.

Novak, Jelena. *Postopera: Reinventing the Voice-Body*. Farnham; Burlington: Ashgate, 2015.

Malani, Nalini. *Transgressions*. Video/shadow play, 2001.

Mutu, Wangechi. *The End of Eating Everything*. Animated video, 2013.

Spivak, Gayatri Chakravorty. "Can the Subaltern Speak?" In *Colonial Discourse and Post-Colonial Theory: A Reader*, edited by Patrick Williams and Laura Chrisman, 66-111. New York: Columbia University Press, 1994.



Culture colors the voice, contours its performative capacities, and leaves deep imprints on its character - it *mediates* the voice, in terms of its accent, intonation, timbre, cadence, and rhythm. And these mediations and their performances matter and are just as powerful as and underpin the electronic effect, be it analog or digital.

(Nori Neumark: *VOICE*)

**You can choose to read and perform the text on the cards simultaneously with others, one after the other, in concert**



## The Collective Vocal Projects

The collective vocal projects *Collective Performative Reading* (2016-) and *[multi'vocal]* (2017-) are presented through a remediation and documentation. Common to the two collective projects is that they are participatory. They invite participants to collectively give voice and listen to and through voices.

### **Collective Performative Reading – Remediation (2020 forthcoming)**

For More Than One Voice. "Kollektiv/Collective," in *Peripeti – Tidsskrift for dramaturgiske studier*. Eds. Cecilie Ullerup Schmidt, Mette Risgård Tranholm, Solveig Daugaard and Stina Hasse Jørgensen. Special Issue, 2020 (forthcoming).

The collective practice-led research project, *Collective Performative Reading*, centers on the politics and poetics of the multiplicity of vocal expression in performance and text, where the complexities of voice are investigated. In the text, we write about the development of our project in relation to both Adriana Cavarero's theory of voices and speech act theory. We then present a remediation of the *Collective Performative Reading* inviting readers as speakers and listeners to do their own collective performative reading, and together investigate how we might speak and listen in more than one voice by means of resonance, polyphony, dissonance, ambiguity, plurality and embrace.

### ***[multi'vocal]*: Reflections on Engaging Everyday People in the Development of a Collective Non-Binary Synthesized Voice – Documentation (2018)**

*Politics of the Machines - Art and After (EVA Copenhagen) conference proceeding*. Eds. Laura Beloff and Morten Søndergaard, 2018.

The collective practice-led research project *[multi'vocal]* investigates how to create multivocal expressions in a synthetic voice as a radical alternative to the existing commercial synthetic voices performing vocal stereotypes of accent, age and gender. In the text we explain the process of collecting voice samples from a diverse range of participants and discuss possible futures developments of the project.





# Collective Performative Reading (2016-)

## CONCEPT

The Collective *For More Than One Voice*  
(Jane Jin Kaisen and Stina Hasse Jørgensen)

## PERFORMANCES

NLH Space, 2016  
Den Frie, 2016  
Sorte Firkant, 2016

## REMEDIATION

“Collective Performative Reading” (artistic remediation)  
*Peripeti*, eds. Cecilie Ullerup Smith, Mette Tranholm, Solveig Dausgaard  
2020 forthcoming.

# Kunstnerisk bidrag

Collective performative reading

# Collective performative reading

*For More Than One Voice*

## The collective *For More Than One Voice*

The voice is interrelational: collective, social and subjective in an ongoing exchange with others who listen and speak. *For More Than One Voice* (2016–) is a collective reading that investigates how we might speak and listen in more than one voice by means of resonance, polyphony, dissonance, ambiguity, plurality and embrace. The collective reading is centred around the politics and poetics of the multiplicity of voices in performance and text where the complexities of voice as representation are investigated.

The collective reading is constituted in three different ways: 1) through collective readings of the politics and poetics of voice in the collective *For More Than One Voice*, 2) through workshops with invited artists concerned with voice and listening for a smaller groups of participants organized by the collective, and 3) through a collective performative reading together with participants in public spaces. In the collective performative reading a selected number of quotes from theoretical, poetical, and political texts on voice and listening are read aloud by participants – who are also invited to listen to the multiplicity of voices unfolding in the spaces. The collective performative reading enables the voices of the participants to destabilize ways in which we think and talk about the voice as representation and a marker of a fixed identity.

## Affinities

*Ci vuole almeno un duetto, un chiamarsi e risponderci: ossia una reciproca intenzione di ascolto, già attiva nell'emissione vocale, che rivela e comunica ognuno all'altro. (Cavarero, 2003)*

*It takes at least a duet, a calling and a responding – or, better, a reciprocal intention to listen, one that is already active in the vocal emission and that reveals and communicates everyone to the other. (Cavarero, 2005)*

The collective performative reading is inspired by philosopher and feminist Adriana Cavarero. In her book *A più voci: Per una filosofia dell'espressione vocale* (2003) Cavarero argues that speech which privileges language and semantic meaning has overruled the sonorous acoustic qualities of the voice. Cavarero argues that we should not only pay attention to what is said, but also to who is saying it and who listens.

In her critique of western metaphysics and the detached philosophical “I”, Cavarero deconstructs the traditional philosophical gender binary, writing that we should reclaim the voice as being unique and plural at the same time. Cavarero argues for a vocalic relationality “where the uniqueness of each existent is constituted and distinguished as a self – a self-in-relation” (Burgess and Murray, 2006) which is a pluralistic and inclusive form of relation.

The English translation of the book *For More than one voice: Toward a Philosophy of Vocal Expression* (2005) gave our collective its name. The collective performative readings emphasize the voice expressing an interrelationality between speaker and listener. We are interested in the voice as a transnational and transhistorical phenomenon, as a collective of the whos-in-relation that speak and listen to each other. We are interested in the voice as an artistic medium unfolding as a linguistic, auditory, bodily, sensuous, and material phenomenon. In our collective work we explore the complexities of voices as phenomena that constitute, destabilize, and push our ideas about who the voice belongs to and the who that listens. We want to shake the notion of voice as a metaphor for one’s identity, as a representation of a singular speaking subject with a fixed identity and subjectivity. We want to disturb our cultural expectations and existing normative understandings of the voice as belonging to a body with only one history, gender, accent, belonging and affiliation. How can we stage a listening of voices destabilizing our social conventions, norms and experiences of the subject as unambiguous and fixed? How can we create spaces that enable the multiplicity, polyphony, multivocality and variability of voices? Inspired by Cavarero we are interested in the relationship between voice and listening, and how listeners actively participate in the creation of meaning in the who of the voice.

*Proprio perché la parola ha una consistenza sonara, parlarsi è comunicarsi nella pluralità delle voci. Detto altrimenti, l'atto del parlare è relazionale: ciò che in esso sempre e prima di tutto si comunica, al di là degli specifici contenuti che le parole comunicano, è la relazionalità acustica, empirica e materiale, delle voci singolari. (Cavarero, 2003)*

*Precisely because speech is sonorous, to speak to one another is to communicate oneself to others in the plurality of voices. In other words, the act of speaking is relational: what it communicates first and foremost, beyond the specific content that the words communicate, is the acoustic, empirical, material relationality of singular voices. (Cavarero, 2005)*

We have invited artists we respect and find inspiring to discuss with us the voice’s potential and actualization in the arts through a series of workshops. The workshops represent a way of creating an ongoing process of inquiry in order to open up artistic practices and theoretical discussions for new ways of forming a dialogue about the voice in art. In our discussions with each other and others in the workshops we have explored what the voice can do as an artistic medium and what listening to voices can do. Here we encompass questions such as: what is the role of the voice today? Who has the right to speak or be silent? And what happens when the voice no longer comes from our bodies but is mediated through technologies or systems? What is the poetic and political nature of the voice? What role does the listener have in relation to the voice?

Listening is never simply a passive, objective and receptive process, but rather an act that plays a

fundamental role in the construction and facilitation of the speech of the interlocutor (whether subject or object). (Abu Hamdan, 2012)

We held the first workshop with Beirut-based artist, Lawrence Abu Hamdan, on the voice of politics, testimony and character. In the workshop with Abu Hamdan we had conversations about the importance of the listener as a powerful interpreter of utterances made by others. With his 2012 article “Aural Contract: Forensic Listening and the Reorganization of the Speaking Subject” Hamdan makes us aware of the role of the listener in relation to the voice. He generates a discussion about how we listen to voices and how this listening connects to broader social and political interests which are also reflected in his artworks such as *The Whole Truth* (2012) and *Conflicted Phonemes* (2012). In the works Hamdan investigates a forensic listening of voice and how this connects to specific historical, social, and political understandings of the speaking subject. We listen to the voice as something that distributes information. But how do we listen to the information the voice transmits? What is the relationship between speaker and listener?

In the workshop by the artist and theorist Brandon LaBelle we discussed the poetic and political potential of the voice.

*- the voice stretches me; it drags me along, as a body bound to politics and poetics, its accents and dialects, its grammars, as well as its handicaps. (LaBelle, 2014)*

Through our work in connection to the workshop with LaBelle we collectively considered the politics connected to the materiality and imaginary of voice when creating spaces for more than one voice. Which voices are listened to and how are they listened to? In the work *Rehearsal for a People's Microphone* (2013) LaBelle considers the mode of political address through the voice of the crowd as an expression of political issues in the public sphere. In LaBelle's book *Lexicon of The Mouth* (2014) we are invited to think of the materiality of the mouth and how the bodily embodiment of speech comes to matter. He points us to the politics and poetics of micro-oralities such as whispering and stuttering, focusing on the materiality of the mouth as a mediator between interior and exterior; he also indicates how the acts of the tongue, lips and throat are connected to social structures enabling or disabling certain ways of making sound with the mouth, voice, speech, and creating an oral imaginary.

### **The collective performative reading**

The collective performative reading has been performed at NLH Space as part of *Stemmer* (2016) curated by Mette Garfield and Miriam Wistreich at Den Frie Udstillingsbygning as part of *Scripted!!* (2016), edited and organized by Trine Mee Sook Gleerup, Maria Bordorff, and Mathias Kryger in collaboration with Eller med a, and at Sorte Firkant as part of the workshop *Translating Geographies of Displacement* Convened by *Migratory Times, the Institute of (im) Possible Subjects, For More Than One Voice*, and at *land's edge* (2016).

## Collective performative reading



*Performance documentation Den Frie Udstillingsbygning, Copenhagen, 2016.  
Photo: Jane Jin Kaisen.*

Who is speaking to you? Who has the space to speak? How are we, collectively, able to give voice? How do you speak? Who is listening to you? How do you listen? What spaces for speaking and listening do you create together with others?

Curator and art theorist Irit Rogoff writes in the article “We – Collectivities, Mutualities, Participations” that “performative collectivity, one that is produced in the very act of being together in the same space and compelled by similar edicts, might just alert us to a form of mutuality which cannot be recognized in the normative modes of shared beliefs, interest or kinship” (Rogoff, 2002). By focusing on the reciprocity of speaking and listening, and by investigating the voice as language and as materiality as well as performative action, the collective performative reading draws attention to the gaps, otherness and multiplicities. It also questions the authenticity of the voice as a representation of a singular fixed subject(ivity). The voice is not only about speaking but also about listening, as the anthropologist Jenny R. Lawy writes:

*Voice (as both political as well as performative) is not only about voicing or speaking; it is also about listening and hearing, for it is in the ways that the audience or listener reacts to, accepts, or rejects what has been put into the social milieu that reveals the (political/social) impact of that voice. (Lawy, 2017)*

The collective performative reading is a performative archive unfolding theoretical, political, and poetic texts on the voice and the voice as a marker of identity. It is performative in that the words read aloud do not only describe the politics and poetics of the voice, they also manifest the voice as political, testimony, and character in relation to the listener. According to philosopher John Langshaw Austin’s speech act theory, words can be perceived as more than something describing a situation. According to Austin, language creates situations through performative sentences, that is, sentences that actually act (Austin, 1976). Or, as performance theorist James Loxley explains, “the utterance is not setting out to *describe* a situation, an event or an action: it *is* an event or an action” (Loxley, 2007). Thus, a statement is not merely a description of what was going on at a given time. The utterance, as a performative speech act, constitutes the event anew. The voice is the medium through which the event takes place. When the utterance reconstitutes the event we must not only relate to *what* is said but also to *how* it is said. The *how* gives new attention to the context in which the speech act is performed and how the vocal performance interacts and relates to the surroundings. It matters whether, for example, a testimony is pronounced in a courtroom or in the street. Or as Austin describes it: if a child says “I do” to their friend acting as if they were getting married the utterance is not a successful performative sentence. It is only when the utterance is said in church or in the town hall, where a priest or mayor can act as a witness, that the utterance becomes a performative sentence as the utterance of the words “I do” results in being married (Austin, 1976). Therefore, speech act theory not only has a focus on the utterance as an action, but also on the social circumstances that can confirm the utterance as an act. It makes the relationship between listening and speaking much more complex. For example, if we consider a witness statement in a court case to be a performative speech act, then it is not only *what* is being said and *how* it is said, but equally the context in which it is uttered (in this case the court room composed of the judge and the jury of listeners) that actualizes the event as a performative sentence.



## Collective performative reading

In this way, the listener's role has changed from being a witness to being a co-producer of events and truth production. The listener brings to the fore the witness not as a potentiality – but as an actuality. The listener's interpretation of what is said and how it is said is crucial to what ends up appearing as the event or action that took place.



*Performance documentation NLH Space, Copenhagen, 2016.*

In the collective performative reading we invite participants to read aloud quotes concerning the voice (as a marker of identity). Participants are invited to actualize the quotes about the voice as performative speech acts, changing the quotes from descriptions of the voice to manifestations of the voice as an event or action; reading the quotes slowly, whispering, growling, overlapping other vocal performances of the written quotes. The participants are also invited to listen and actively witness the voice spoken as action and event by others in the room. As such, we ask participants to enable the performative speech act by collectively



shifting positions between reading aloud quotes about the voice as performative sentences, and listening to and confirming the voices heard as performative sentences. In this way participants enter into a committed relationship between speaking and listening to performative sentences, creating voices as events that unfold across time and space. Participants are invited to listen to the voices of others as events: voices that merge, resonate, dissonate, fluctuate, build upon each other, clash together, intersect, and give space and new meaning to each other. Participants are invited to change positions in the room and give each other a place from which to speak and be heard. We invite the participants to listen to broken pieces of words in Danish, English, French, German, Farsi and Korean that emerge and disappear into a chorus of the different voices of the participants. To listen to different languages that dissolve and become affective sound surfaces, rhythms, and fluid materiality that mingle moving back and forth between participants in the room. We often focus on the voice as speech, which creates meaning through language but in the collective performative reading we ask participants to also focus on the reverberation and texture of voices. We ask the participants, as listeners, to pay attention to the various sound registers of the voices, an emotional resonance, tonality, rhythm, pauses of speech – the performative elements of the utterance of words when they hear the chorus of voices created by the other participants. The participants are invited to listen to the theoretical, political, and poetical descriptions of the voice as well as how these descriptions are uttered affectively.

Through the duet of speaking and listening the voices become events that change from moment to moment. Participants are invited to listen to how words are performed by the voice. To listen to the voice as a material practice and as a metaphor that unfolds and gets its meaning in its relation to the listener as something continuously changing, fluctuating and unfolding in the interrelationality. The participants form a chorus of diverse voices pointing beyond the individual to a collective event of speakers and listeners with the potential to actualize the performative sentences being spoken and heard. It is a collectivity that opens to the complexities of poetic, political, and cultural affinities. By focusing on how a statement is expressed and heard, and how the context of the statement affects the experience of what happened and how this is represented, the collective performative reading also enables a focus on the politics of representation. By working with how we listen to the voices of collective performative reading we can identify and complicate representations. We can enable multiplicity, otherness, and a productive questioning of unidirectional representations through the use of voice in art - and how we listen to it. In the collective performative reading we ask: What voices have the right to speak and be heard as performative sentence? Which listeners have the power and position to confirm the vocalization of the readings as performative sentences, making the reading into an event and action instead of a description?

## Collective performative reading



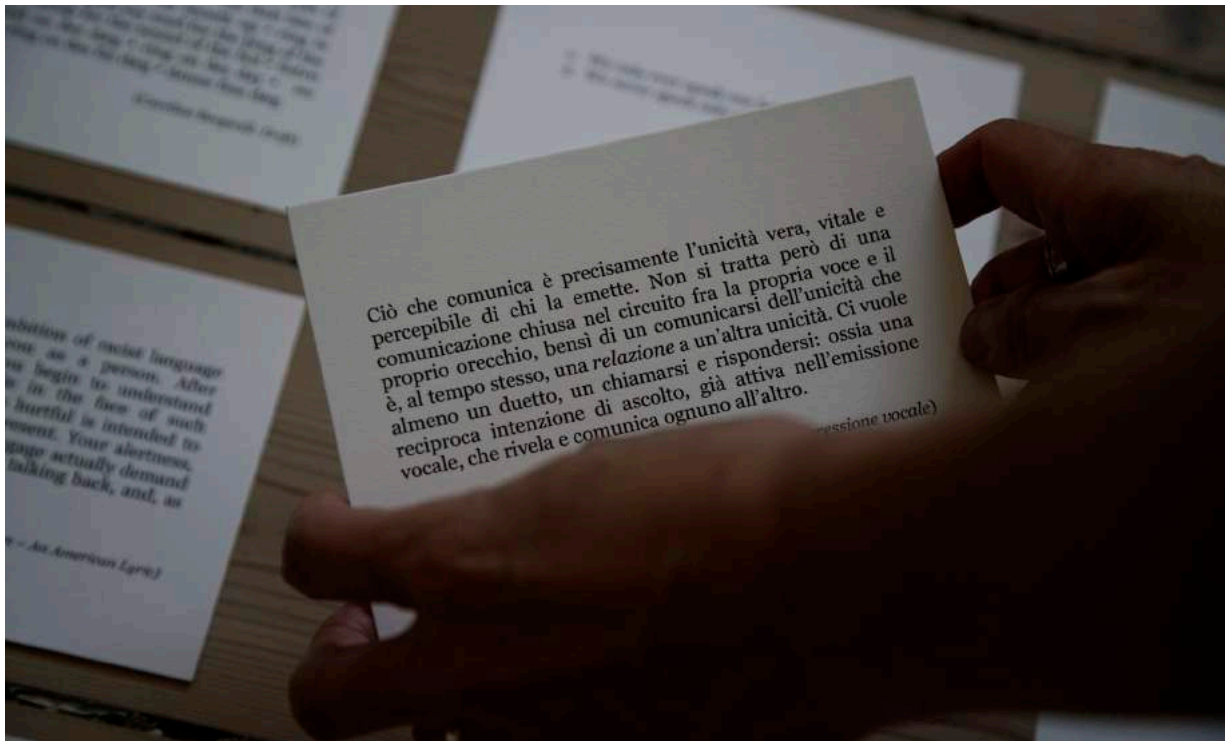
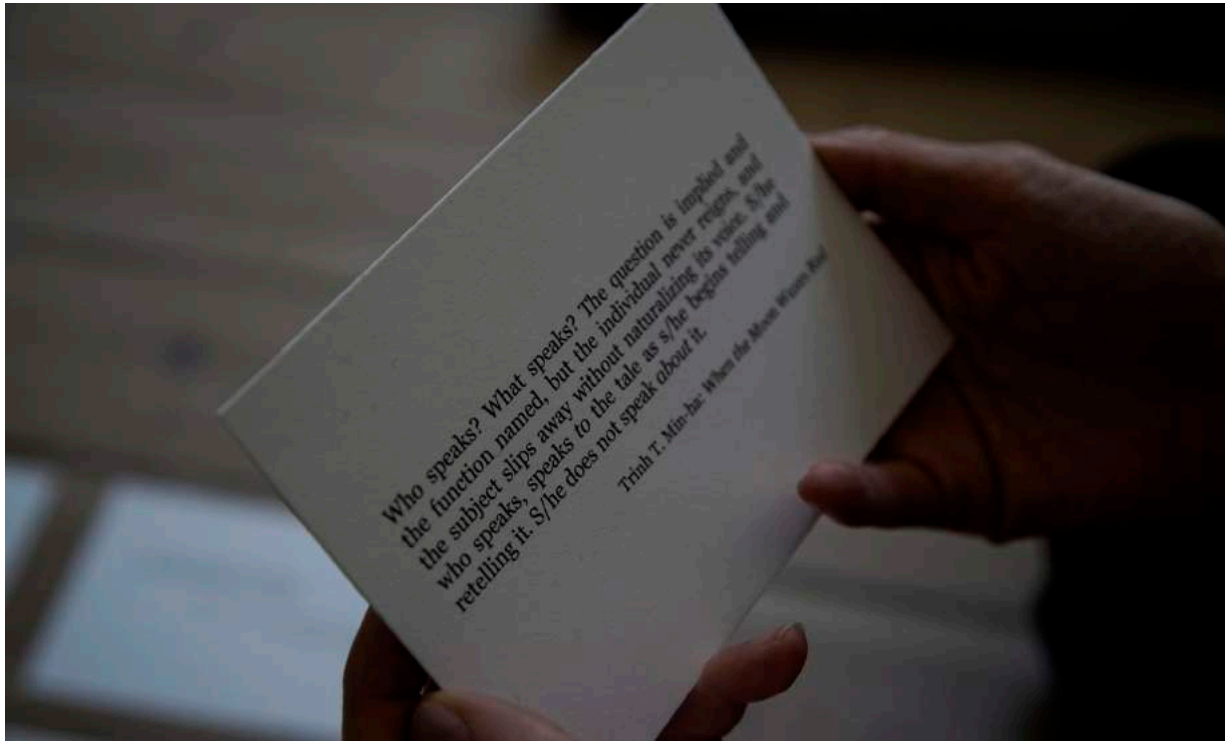
*Performance documentation Sorte Firkant, Copenhagen, 2016. Photo: Jane Jin Kaisen and Stina Hasse Jørgensen.*

## Guide to a collective performative reading

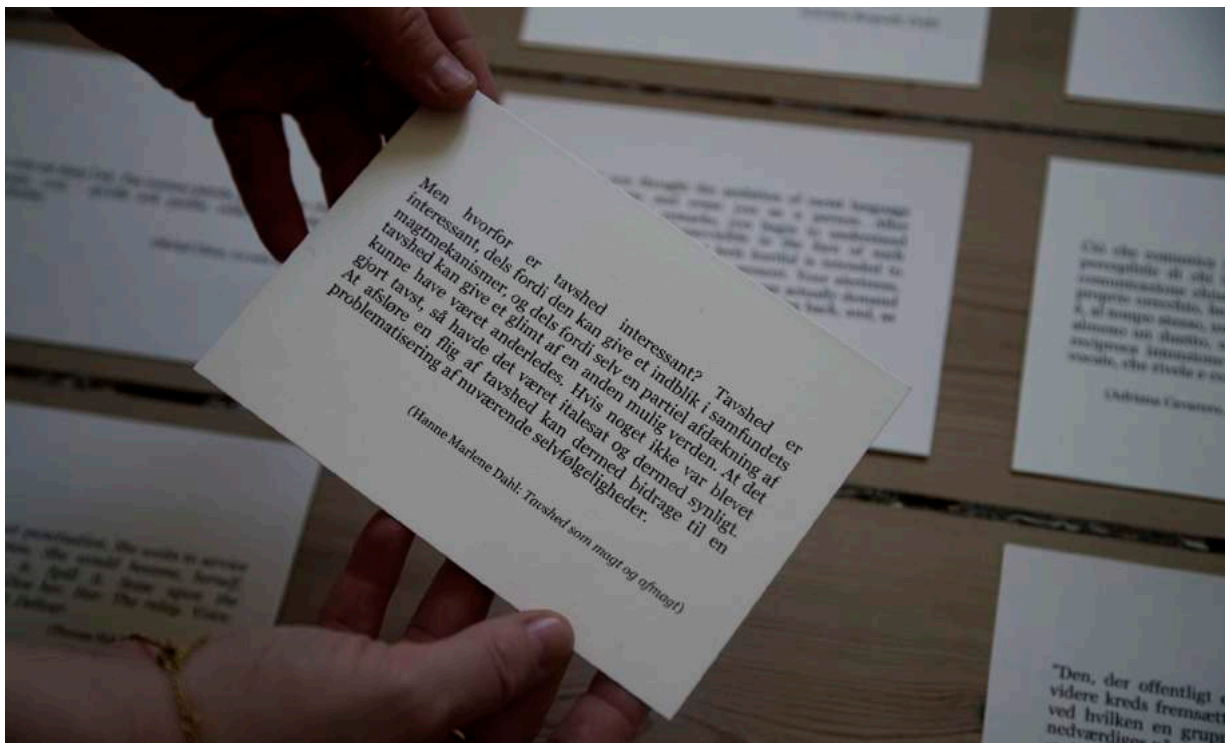
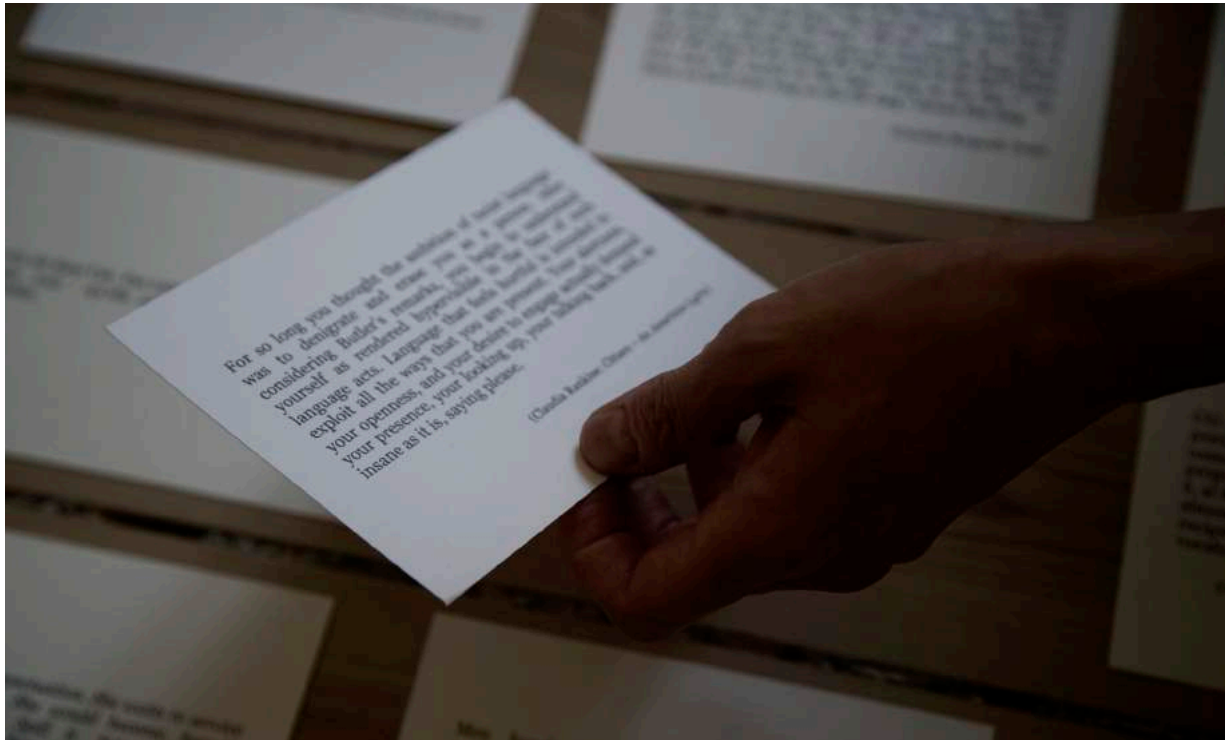
*(A series of cards with quotes, statements, definitions, and reflections on Voice written in different languages are distributed on the floor. They can be placed in any desired configuration.)*

(Speaker): *For More Than One Voice* is a collective performative reading that investigates the relationship between the written and the spoken word. Specifically, how we can read, perform and listen to *more than one voice* in ways that allow resonance, polyphony, dissonance, ambiguity, plurality, and embrace.

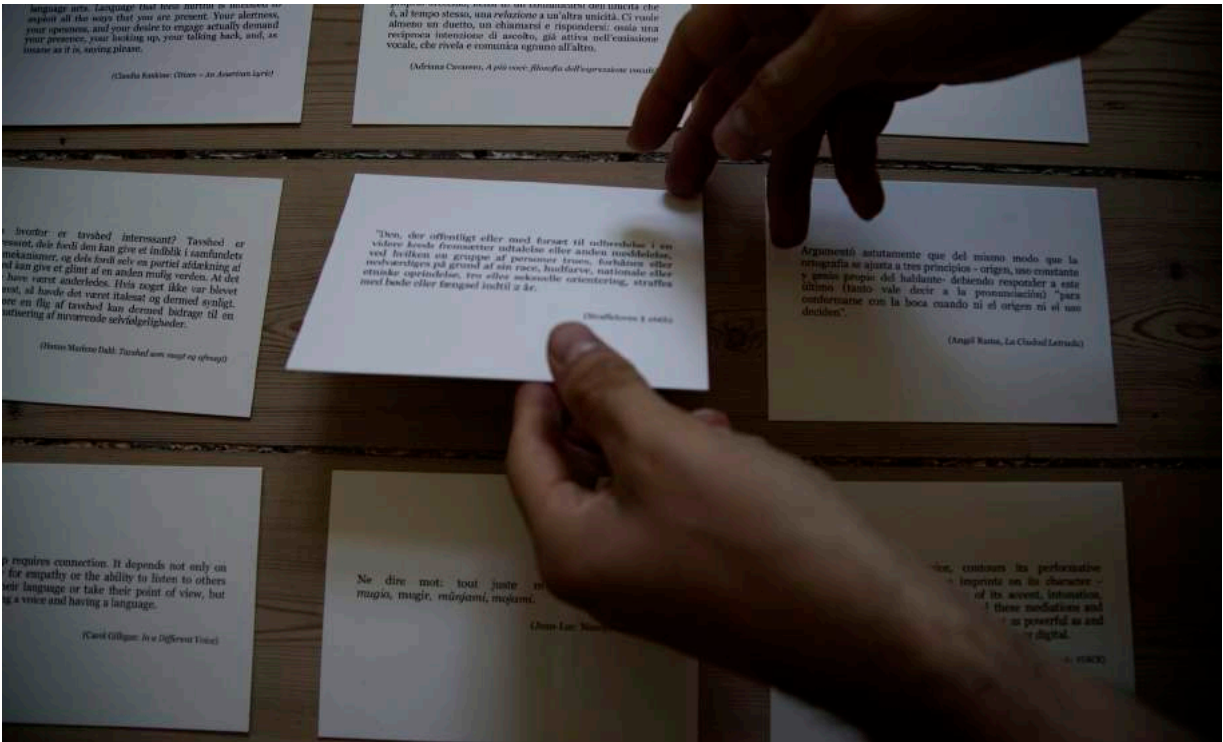
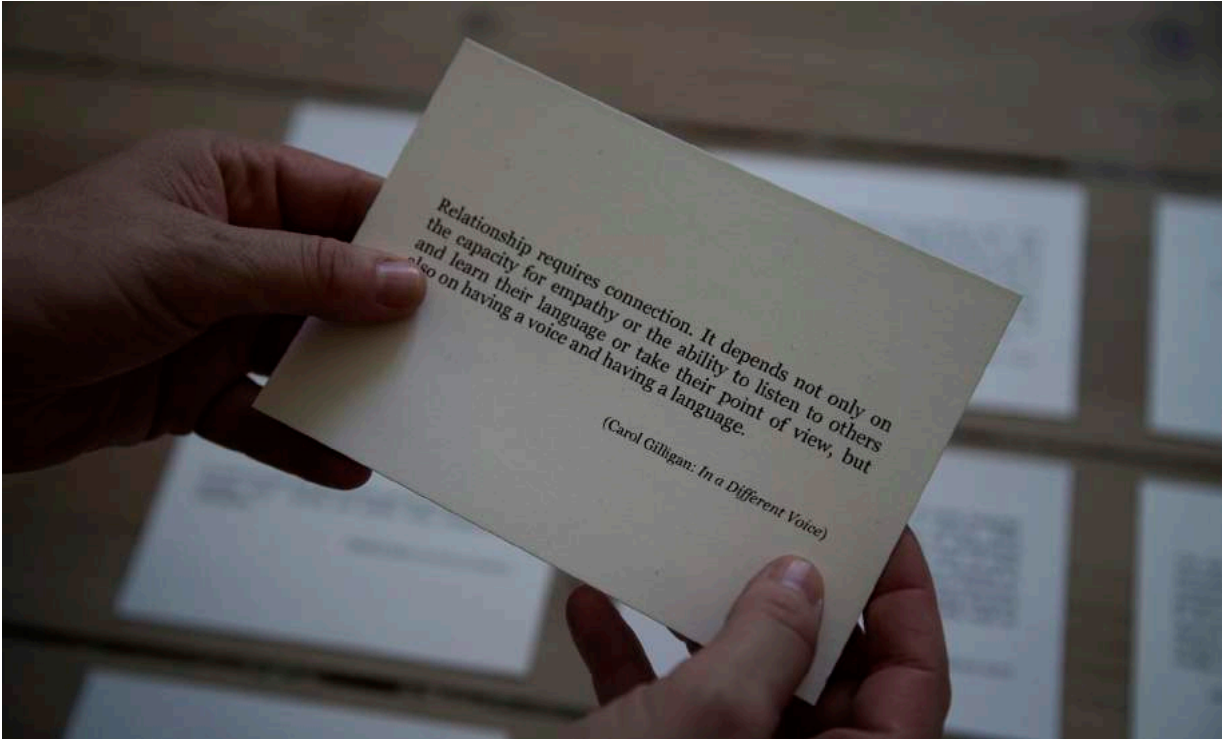
(Another speaker): We invite you to read the quotes on the cards (or what you are able to read of them – the rest you have to voice in your own way). You are welcome to cut out the cards and place them on a floor or a table. Then you can choose and read out a quote at any time and at your own pace. You can shout, sing, yell, scream, murmur, mumble, or whisper the words you read. You can stick out your tongue and think about the feeling it installs. You can gasp, cough, burb, grunt, (whistle) sigh, pause, or yawn as you like. You can choose to read the text on the cards simultaneously, one after the other. You can listen to your collective readings together. Hear how your voices resonate, dissonate, fluctuate, oscillate, overpower, overlap. You may keep the quote, pass it on to another participant or put it back on the floor or table for someone else to possibly read. You can read more than one quote. We also invite you to take a quote that you are unable to read or pronounce. Think about how this makes you feel. Please also feel free to listen and be actively silent. Once you hear that all the voices have come to rest the collective performative reading has come to an end.

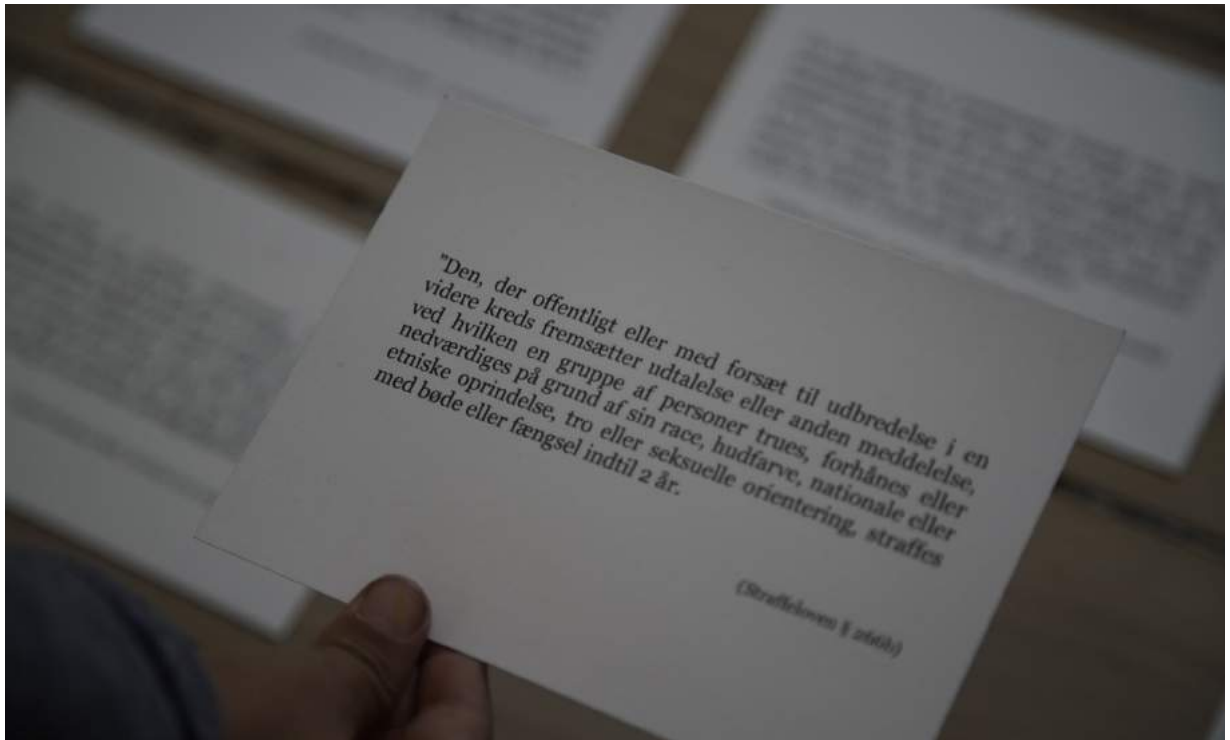




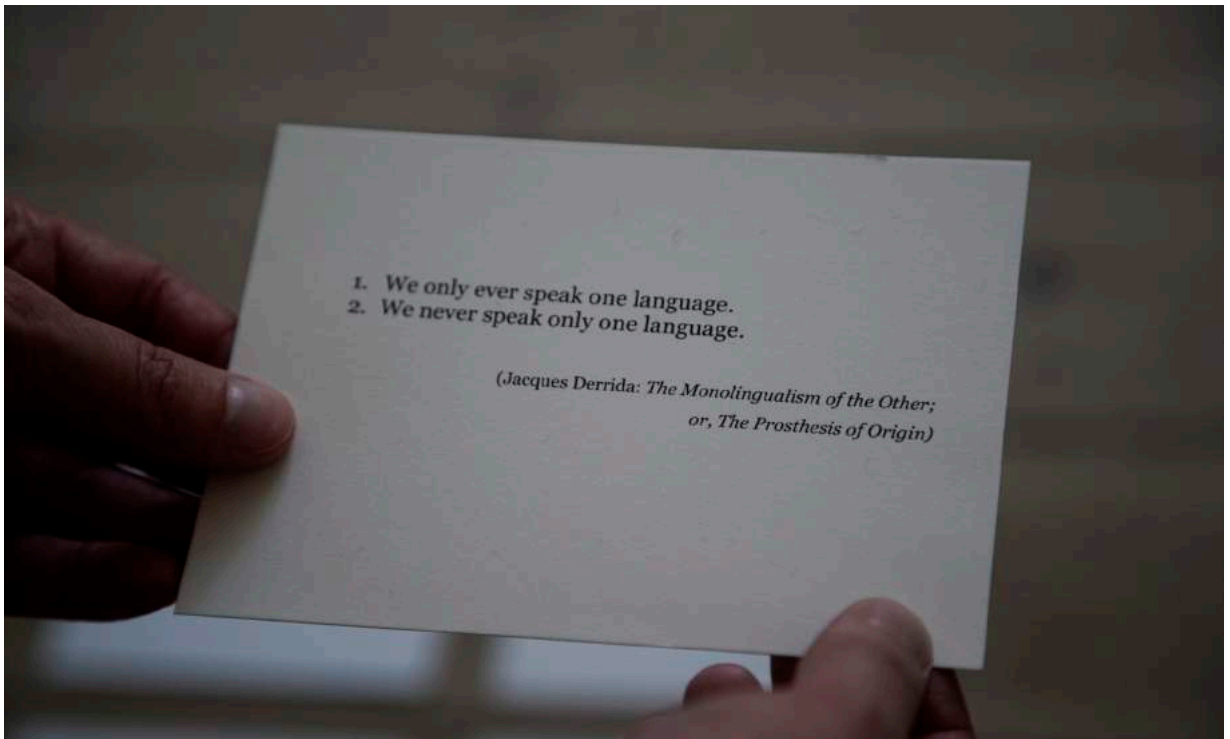
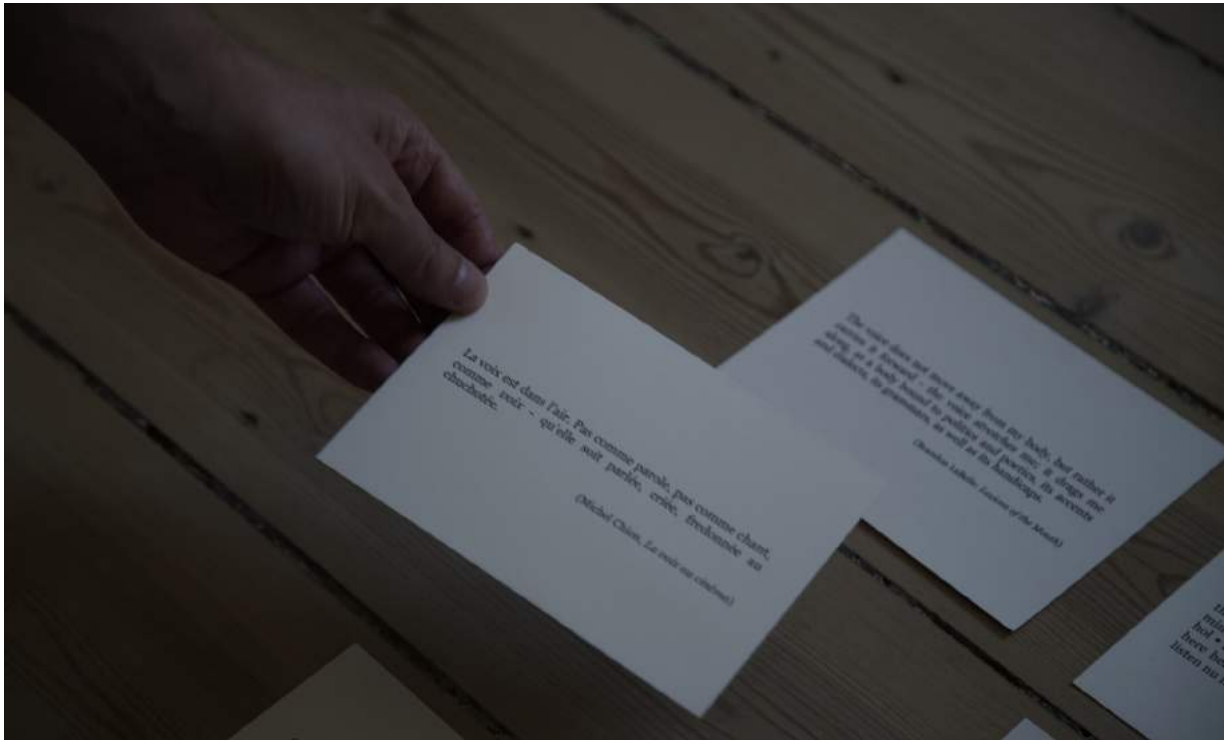


# Collective performative reading





## Collective performative reading





## Abstract

*For More Than One Voice* (2016–) is a collective reading that investigates how we might speak and listen in more than one voice by means of resonance, polyphony, dissonance, ambiguity, plurality, and embrace. The collective reading is centred around the politics and poetics of the multiplicity of voices in performance and text, where the complexities of voice as representation are investigated.

## Abstrakt

*For More Than One Voice* (2016–) er en kollektiv læsning der undersøger hvordan vi taler og lytter med mere end en stemme gennem resonans, polyfoni, dissonans, tvetydighed, pluralitet og omfavelse. Den kollektive læsning er centreret omkring flerstemmighedens politik og æstetik i performance og tekst, hvor stemmens kompleksiteter som repræsentation er undersøgt.

## References

- Austin, J. L. 1976. *How to Do Things with Words*. Oxford: Clarendon Press.
- Burgess, S.K., Murray S.J., 2006. For More than One Voice: Toward a Philosophy of Vocal Expression (review), *Philosophy & Rhetoric*, Vol. 39 (2), pp. 166–169.
- Cavarero, A., 2003. *A più voci: Per una filosofia dell'espressione vocale*. Milano: Feltrinelli Editore.
- Cavarero, A., 2005. *For more than one voice: Toward a Philosophy of Vocal Expression*. Stanford: Stanford University Press.
- Abu Hamdan, L., 2012. *Aural Contract: Forensic Listening and the Reorganization of the Speaking Subject*, *Cesura//Acceso*.
- Jørgensen, S. H., 2019. Stemmernes politik i samtidskunsten. In: eds. C. J. Jepsen, R. Kjærboe, S. Krogh og M. Søberg. *Terræn. Veje ind i samtidskunsten*. Aarhus: Aarhus Universitetsforlag, pp. 228–245.
- LaBelle, B., 2014. *Lexicon of The Mouth*. London: Bloomsbury Publishing.
- Lawy, J., 2017. Theorizing voice: Performativity, politics and listening. *Anthropological Theory*, Vol. 17 (2), pp.192–215.
- Loxley, J., 2007. *Performativity*. London: Routledge.
- Rogoff, I., 2002. 'We - Mutualities, Collectivities, Participations. In: ed. *I Promise its Political*. Museum Ludwig: Cologne, pp.126–134.





## [multi'vocal] (2017-)

### CONCEPT

The *[multi'vocal]* collective

(Alice Emily Baird, Frederik Tollund Juutilainen, Mads Pelt and Stina Hasse Jørgensen)

### INSTALLATIONS

Roskilde Festival, 2017

Tech Festival, 2017

IT University of Copenhagen, 2017

The Danish Society of Engineers, 2017

Catch at Click Festival, 2018

Catch Exhibition, 2018-2020

### RELEASE

Anyines Label, 2020

### DOCUMENTATION

“*[multi'vocal]*: Reflections on Engaging Everyday People in the Development of a Collective Non-Binary Synthesized Voice”

*Proceedings of Politics of the Machine*, The British Computer Society,  
edited by Morten Søndergaard and Laura Beloff.

Co-authored with Alice Emily Baird, Frederik Tollund Juutilainen, Mads Pelt and Nina  
Cecilie Højholdt

# [multi'vocal]: Reflections on Engaging Everyday People in the Development of a Collective Non-Binary Synthesized Voice

Stina Hasse Jørgensen  
University of Copenhagen, Department  
of Arts and Cultural Studies  
Karen Blixens Vej 1, 2300  
Copenhagen S, Denmark  
[rhj282@hum.ku.dk](mailto:rhj282@hum.ku.dk)

Alice Emily Baird  
Augsburg University, Chair of  
Embedded Intelligence for Health Care  
and Wellbeing, Eichleitnerstr. 30, F2,  
86159 Augsburg, Germany  
[alice.baird@informatik.uni-augsburg.de](mailto:alice.baird@informatik.uni-augsburg.de)

Frederik Tollund Juutilainen  
University of Copenhagen, Department  
of Nordic Studies and Linguistics,  
Karen Blixens Vej 1, 2300  
Copenhagen S, Denmark  
[frederik@tollund.cc](mailto:frederik@tollund.cc)

Mads Pelt  
Netcompany, Grønningen 17, 1270  
København K, Denmark  
[madspelt@gmail.com](mailto:madspelt@gmail.com)

Nina Cecilie Højholdt  
Malmö University, K3 Arts and  
Communication, Interaction Design,  
Östra Vervsgatan 11A, 211 19 Malmö,  
Sweden  
[nhoejholdt@gmail.com](mailto:nhoejholdt@gmail.com)

**The growing field of Human-Computer Interaction (HCI) takes a step out from conventional screen-based interactions, creating new scenarios, in which voice synthesis and voice recognition become important elements. Such voices are commonly created through concatenative or parametric synthesis methods, which access large voice corpora, pre-recorded by a single professional voice actor. These designed voices arguably propagate representations of gender binary identities. In this paper we present our project, [multi'vocal], which aims to challenge the current gender binary representations in synthesized voices. More specifically we explore if it is possible to create a non-binary synthesized voice through engaging everyday people of diverse backgrounds in giving voice to a collective synthesized voice of all genders, ages and accents.**

*Engagement. Non-binary. Identification. Synthesized voices. Interaction Design. Corpus collection.*

## 1. INTRODUCTION

Giving voice to machines as a means of improving human-computer interaction, is a tool continually implemented by many large technology companies. Synthesized voices are now integrated into a vast array of daily interactions, such as smart-devices assistances (i. e. Apple's®, Siri®), public transport announcers, and enhancing humanoid robot identity (i. e. SoftBank Robotics, NAO). Current synthesized voices add a human-like personality to the machine, to improve human-computer-interactions (HCI) (Clifford & Kwan Min 2000).

The ability to computationally recreate the human voice has a long history in the field of engineering (Schroeder 1966, Young 1979). Conventional signal processing methods such as phoneme synthesis (Gagnon 1978), were for a period, the

state-of-the-art in the field. Although segmentation was proposed much earlier (Peterson, Wang & Sivertsen 1958), it was towards the end of the 1980's that speech synthesis moved from signal replication, to computationally efficient methods for managing recorded human speech corpora, a method commonly known as concatenative voice synthesis (Hunt & Black 1996).

Methods for speech synthesis production now include parametric systems which run through trained Deep Neural Networks (DNN) (Zen, Senior & Schuster 2013), and require a large speech corpus, recorded by a single voice. The voice is recorded by a single professional voice actor, in laboratory conditions, and now-a-days may even include varying emotional states, to allow for diverse speaking styles (Iida et al. 2003). Although diversity is being pushed in terms of speaking style,

to avoid the stereotypical monotonous robot-like speech, there is little attention being paid to the consequences that result from the current skeuomorphic synthesized vocal identities being created through these methods.

There is a need to study and discuss the politics of the digitally designed voices in the era of algorithmic computation, especially in relation to what representation and identities are evoked through the increasing prevalence of synthesized voices in everyday life, as well as to explore alternative representational strategies to the existing synthesized voices.

Here we will present our project, which is driven by the question; Since machines are not limited to a single identity, why do machine voices have only one gender, one age, and one accent? As an alternative to the existing skeuomorphic synthesized voices, we seek to develop a collective synthesized voice enabling participation in corpus design, and engaging different everyday people in the development process in terms of open-source frameworks. Here it is important to point out that we do not believe that one participant necessarily equals one gender, but that this participant can represent multiple genders as we do not conceive of gender as something essentialist or static, but as fluid and temporally unfolded (Jones 2012).

In the following a short presentation of voice synthesis technology will be given, followed by an introduction to the feminist and queer theories of gender that have motivated our research and design project. Then we present our project, [multi'vocal], and our methodological approach with a specific focus on the design choices and principles implemented in order to create a collective synthesized voice, by enabling everyday people to engage in the corpus collection and open-source frameworks for a gender non-binary collective synthesized voice. Our project will then briefly be contextualized through an introduction to related work in the field. Finally, a short discussion of the challenges that emerge when designing for creating a collective non-binary synthesized voice are made, followed by a conclusion and a reflection on future work possibilities.

## **2. LITERATURE REVIEW: NON-BINARY GENDER AND SYNTHESIZED VOICES**

### **2.1 Synthesized voices**

Text-to-speech (TTS) systems are a common way to generate synthesized speech, allowing users to freely input their desired text. Such systems are often integrating adaptations of Time-Domain Pitch Synchronous OverLap- Add (TD-PSOA) based,

concatenative synthesis (Dutoit & Leich 1993), a method which requires a large corpus of pre-recorded human speech, segmented into phoneme sized instances, allocated from a text input, forming synthesized replications of the sentences (Boughazi & Tabet 2011). Parametric speech synthesis, also used in TTS systems, has the addition of DNN's, to computationally improve the selection process, allowing for more human-like features (Barra-Chicote et al. 2010). Additionally, some DNN based frameworks, such as WaveNet (Zen et al. 2016), are now challenging the state-of-the-art for speech synthesis.

With corpus collection being the driver for such systems, a crucial element for engineers during this stage is to achieve consistency across the entire corpus, both for pronunciation and phrasing (Saratxaga et al. 2006). This is important in order to achieve intelligibility, as segments may be taken from differing instances, depending on the text input (FestVox 2014a). In order to achieve such qualities recordings are made in laboratory conditions or studios, with recordings for a corpus size with 3000+ sentences (FestVox 2014b). These digitally designed voices are often categorized in terms of language and the gender represented (Baird et al. 2017).

### **2.2 Non-binary Gender**

Following the philosopher and queer theorist Judith Butler the gender binary is part of a regulatory practice that upholds the heterosexual hegemony through acts that reiterate the binary of male and female. These normative constraints produce a regulation of who comes to matter as socially visible subjects (Butler 1993).

As it is argued in Baird et al. (2017) and Phan (2017) synthesized voices such as Apple's Siri or the 13 different voices in the IBM Watson archive are categorised within a binary system as either male or female. The politics of digitally designed voices in the era of algorithmic computation lies among other things, in the way in which these voices are categorized within a gender binary system. Because this can be argued to demonstrate a normative understanding of gender within a strict binary frame, where the synthesized voices are rubricated in a classification system that divides gender into two distinct categories of male and female, regarded as oppositional and disconnected. In a study by Baird et al. (2018) however, the experience of gender in synthesized voices are shown not to be merely a binary decision by listeners. This is an important observation that challenges the state-of-the-art for how researchers and companies classify and communicate the identity constructions of synthesized voices.

In the following we will present how theories of the non-binary gender identifications have inspired our development of the project, [multi'vocal], then we describe how specific focus have guided our design choices and the principles implemented in order to engage everyday people in the corpus collection and open-source frameworks for gender non-binary collective synthesized voice.

### 3. DESCRIPTION OF PROJECT: [MULTI'VOCAL]

In this section we present our project, [multi'vocal], and how it has been inspired by queer feminist thinking. Then we describe how the project has been designed in relation to the technical specifications required when creating an open-source synthesized voice. Here we specifically present our methodological approach in the quest to create a non-binary gender synthesized voice by inviting people with all genders, ages and accents to contribute their voice to the corpus collection. We describe how this strategy is supported through our design and principles of how to create the technical system. From a technical perspective our design, research and art collective (henceforth mv) aims to create a technology enabling an infinitely expanding synthesized voice, composed from an increasing corpus of diverse vocal identities. This section provides an introduction to how queer feminist thinking have inspired our project [multi'vocal], the design and technical framework, and finally to related work.

#### 3.1. [multi'vocal]: non-binary synthesized voice

Our project, [multi'vocal], can be defined as a feminist activist project that explores how synthesized voices can be named, categorised and produced differently. In our project we seek to challenge the gender binary categorisation and production of synthesized voices, by creating a non-binary voice where all genders (ages and accents) presented by the voices from participants are heard. The non-binary is a category for gender identities that are outside or not solely male or female, not conforming to the norms of the gender binary. Our aim is to push the ways in which we think, talk and frame representation and identity constructions in technologies such as synthesized voices. As an alternative to the existing synthesized voices, we want to create a synthesized voice with a proliferation and multiplication of gender categories, not static but as fluid and temporally unfolding categories. Through this we want to explore if it is possible to create a representational technology that enhances identification as process, complicating fixed binary identities; a non-binary voice that moves beyond the fixed binary identity.

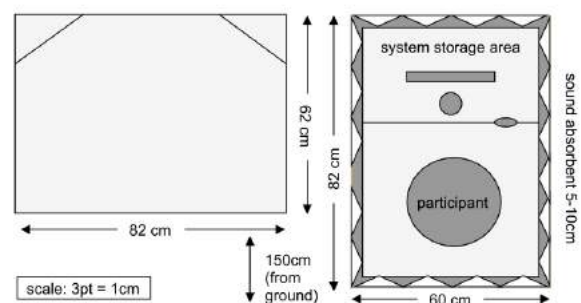
Identification as process is a term discussed by cultural theorist Stuart Hall, who wrote "Perhaps instead of thinking of identity as an already accomplished fact, which the new cultural practices then represent, we should think, instead, of identity as a 'production' which is never complete, always in process, and always constituted within, not outside, representation" (1994 p. 392). Art history scholar Amelia Jones has explained identification as process as "how subjects might navigate the world through process rather than endless oppositional projections that seek to fix others in place in order to confirm the self" (2012 p. 229).

With our project, [multi'vocal] we aim to create a synthesized voice that enhance identification as process, by inviting many different people with all genders, accents and ages to contribute their voice to the corpus collection and design of the synthesized voice. This we hope will complicate the fixed binary identities that are currently dominating the field of voice synthesis.

#### 3.2 Design and technical frameworks

mv has prototyped a design of a data acquisition recording system (henceforth: [multi'vocal] system), which deviates from conventional corporate methods as presented in the 'Introduction', encouraging both participation and discussion in the field of voice synthesis, engaging the participants in the project in public places, events and on online open-source fora such as github.

The design choices and principles implemented by mv, are in order to enable the engagement and participation of everyday people in the corpus collection, and open-source frameworks for collective synthesized voices are presented in the following.



**Figure 1:** The dimensions of the [multi'vocal] system, from side (left), and above (right) views. Materials and equipment were left at the bare minimum to ensure an easily replicable design, and protect against unattended conditions.



### 3.2.1. Design

When developing the physical structure priorities for the [multi'vocal] system have included robustness, affordability and modularity; so that the level in which individual components of the [multi'vocal] system is possible to reconstruct for everyday people around (most of) the world. Dimension details of the [multi'vocal] system are shown in Figure 1.



**Figure 2:** The [multi'vocal] system in use. Different people from all over the world are able to enter the system whilst standing, and contribute their voice to the [multi'vocal] synthesized voice corpus collected at international festivals and events.

The physical structure consists of a head height suspended container (Cf. Figure 2) open at the bottom to allow participants to enter. The structure is placed at height covering to at least the shoulders (blocking peripheral view), with a footstool available for differing heights. The exterior of the [multi'vocal] system is made entirely from plywood, in order to keep the overall mass lightweight, whilst retaining robustness, given the types of locations it will be positioned in. Underneath the wooden exterior 5cm of sound insulation material is fixed, in order to reduce the external background noise.

When participants enter the [multi'vocal] system the screen itself is placed on the back wall, and around average eyes level. The microphone is directly below this, and the button to the right is fixed to an acrylic glass screen, used to protect all elements.

### 3.2.2. Technical frameworks

For the front-end much like reasons outlined in the Section 'Design', robustness, modularity and affordability were prioritised, so off-the-shelf equipment were chosen. The full list of front-end equipment (minus cabling) is as follows: Dynamic microphone (Shure SM57), 11" monitor display, 2

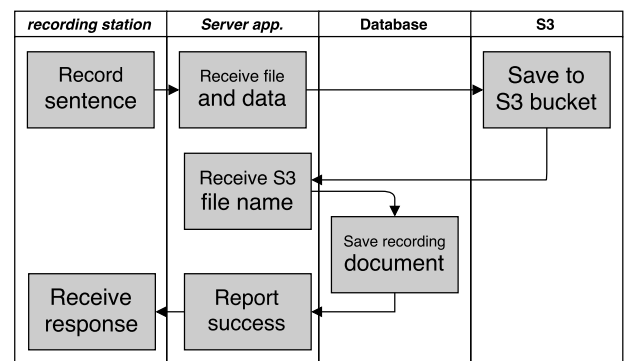
channel audio interface (M-Audio Fasttrack Pro), Single-board computer (Raspberry Pi), Microcontroller (Arduino) connected to Push Button.

At its core mv aims to make modularity of frameworks possible, contributing to the open-source community and providing full documentation in each stage. Both front and back-end code is freely available through the mv github.



**Figure 3:** Simplistic, Graphical User Interface design, implemented inside the recording system. Showing an example sentence read aloud by participants. A full data set of phonetically and prosodically balanced sentence transcriptions is available from (FestVox 2014c).

At run-time the GUI shows a white screen with black and grey text (Cf. Figure 3), a minimalist choice to avoid overloading the view of the participant. Simplicity in the design and experience is aimed at in order to invite as many different participants as possible (Feinberg & Murphy 2000). When the participant holds the record push button, the recording begins and on release the recording ends. The resulting wav file (sample rate of 44.1kHz, and bit rate of 16), is then added to an offline uploading queue and the next sentence transcription is immediately loaded to the display. Since Internet was not always secure, this approach ensured that recordings were not lost.



**Figure 4:** [multi'vocal] system back-end processing, from upload to storage, applied to each recorded speech instances.



The [multi'vocal] system runs all back-end (Cf. Figure 4.) components on Amazon Web Services(AWS) (Amazon 2017a); recordings are saved directly onto a Simple Storage Service (S3) (Amazon 2017b) bucket, and the recording application is running on an Elastic Cloud Computer (EC2) (Amazon 2017c) instance. Mass storage in S3 is inexpensive, enabling mv to indiscriminately save all received recordings.

Verification and management of the continually increasing corpus runs on one of the smallest possible instances on AWS, allowing for the [multi'vocal] system to run inexpensively. When the application receives a recording, the file is saved to the s3 bucket, inserting a document, which includes the name of the file and metadata. The metadata consists of;

- client identifier (ID) - a unique ID for that recording instance, as well as a location ID, so that data collected from differing events or sources can later be managed.
- transcription ID - to indicate the transcription related to the recorded instance.

No other metadata is collected on the individual recordings. However, since all contain the client ID data this still allows mv to differentiate between recordings to produce varying synthesized voice corpus compositions.

### **3.3 Related Work**

Data collection, specifically for speech informs many Interactive Systems, including TTS and Automatic Speech Recognition (ASR). Predominately online, there are projects now exploring different avenues of this. Although to the best of the authors knowledge there are no projects working on diverse corpus construction, specifically for the synthesized voice, aside from [multi'vocal] and the Mozilla Corporation, a non-profit open source technology giant, that is now working on the project Common Voice (Mozilla Corporation 2017) an online collaborative data collection system which gathers speech recordings from diverse people across the world to contribute to an open source ASR system. With speech recognition being limited to the corpus it is trained on, this project is the other side of the coin to [multi'vocal], aiming to improve diversity specifically in ASR, removing the limitation of current state-of-the-art corpora.

## **4. ENGAGEMENT OF EVERYDAY PEOPLE IN THE [MULTI'VOCAL] VOICE CORPUS COLLECTION**

Here we will briefly discuss our experiences so far on how to engage and include the diversity of

everyday people on festivals and public events in the production of the collective gender non-binary synthesized voice, [multi'vocal].

The engagement of everyday people in the development of a design are used commonly today within the field of HCI (Huybrechts 2014), but not in speech synthesis. Yet consistently voice synthesis corpora are created with blackboxed technologies and the voice corpus are gathered in controlled environments (FestVox 2014b) with one professional voice actor. In this way the development of state-of-the-art synthesized voices offers limited agency to everyday people in terms of technological transparency and identity adaptation - something which in regards to e.g. gender identification, may show to have societal consequences (Phan 2017).

The [multi'vocal] project aims to open up the field of voice synthesis to more diverse participation, enabling engagement from everyday people in the open-source frameworks hosted online, and through housing the [multi'vocal] system in environments outside of laboratory conditions, to create a collaborative synthesized voice by engaging everyday people from all over the world. So far utilising public festivals and events such as Roskilde Festival 2017, Techfestival 2017, and Catch at CLICK festival 2018 (now it the project is installed at Catch an art and innovation space, where it will stay until end of 2018), for speech corpora collection, has insured an inclusion of individuals with diverse backgrounds, across a short period of time. Typically, the recordings for the speech corpora are made in a controlled environment, with somewhat strict direction from engineers. mv's approach is more spontaneous (Ward 1989), where the Graphical User Interface (GUI) design (Cf. Figure 2) provides the only guidance for the person recording their voice as part of the voice corpus collection to the collective synthesized voice. However, the lack of strict guidance in relation to the engagement of everyday people in the voice corpus collection has sometimes resulted in unusable recordings with background noise or recordings of half sentences, which brings a workload to manually sorting the voice recordings.

As [multi'vocal] is political project we have also made several talks where the feminist activist agenda is presented with the aim of sparking debate of representation in current technologies.

An ongoing challenge is still however, how we can get beyond the notion of fixed categories when we are still using categories such as gender, age and accent in order to talk about the identity constructions and representations we want to explore through our project. Can we use the

categories that we critique and want to get beyond? Can data-driven systems that operate with annotations and archival categories create a place for identification at all? Or do we, by operating with categories such as gender, age and accent reiterate the norms and notions of fixed binaries? We have decided not to ask participants about their gender, age or accent when they contribute with their voice in order to avoid using fixed categories as little as possible. The problem of technologies operating and reiterating fixed gender binaries, however, need more work in the future.

## 5. CONCLUSION AND FUTURE WORK

In this paper we have presented [multi'vocal], a project exploring if the chosen methodology can engage and include the diversity of everyday people in the creation of a non-binary synthesized voice.

Here we have first presented the influence from feminist and queer theory discussing possibilities of creating synthesized voices beyond the gender binary enhancing identification as process instead of fixed identity constructions. Then we presented the methodological and technical considerations in terms of design choices and data collection principles, followed by a short discussion of the challenges that emerge when designing for the engagement of everyday people, especially in regards to the development of a collective non-binary synthesized voice.

One of the challenges in this project is how we make the exploration of the alternative synthesized voices available to diverse group of people with different backgrounds. The aim of the project is to include voices from people with different genders, ages and accents, yet so far we have only been able to engage people on-site at different festivals and events in Denmark. Through the development of an online platform we hope to engage with people internationally, yet we are aware that this platform will only be accessible to people who are able to go online excluding millions of people worldwide from participating in the project.

Moving forward the methodology of the [multi'vocal] project will be refined in terms of enabling everyday people to participate in the voice corpus collection online, so people not being physically present where the [multi'vocal] project is installed, are still able to contribute. Also refining the data collection methodologies will be an area for advancement in terms of noise reduction and sound quality in general. At a later point an online annotation protocol to the [multi'vocal] system back-end, storing information and recording quality, should be developed. Another future development will be to

use the knowledge gained from making the [multi'vocal] project in the English language, to enable participation of everyday people in the development of synthesized voices in other languages as well. This goal also connect to the overall aim for the mv collective, creating a methodology that can enable engagement of everyday people in the development of a collective non-binary synthesized voices, adding to the diversity of synthesized voices in all languages.

We are also concerned with whether the everyday people participating in the project on festivals and events are aware of the aim with [multi'vocal]; that our project is a feminist activist project that should spark discussion and make participants reflect upon if and how they want identity constructions and representations in technologies to be now and in the future. We have communicated our project aims in written text making the aim explicit. This text is shown next to the voice booth and on websites of the festivals and events where the project is exhibited and presented. In the future the text will possibly stress the fact that we do not believe that one participant equals one gender, but that gender is performative and temporal and can change over time. Due to this a participant is invited to contribute as many times as they wish, expressing multiple identifications.

We hope that we will be able to make a mobile [multi'vocal] voice recording booth, so that we can engage everyday people in other countries and in places where people are not so socially or physically mobile, in order to collect voices to the [multi'vocal] speech corpora.

## 6. ACKNOWLEDGEMENTS

We thank all the participants who have contributed with their voices on festivals and venues, and all the mv collective friends without whom [multi'vocal] would never have been possible.

## 7. REFERENCES

- Amazon. 2017a. Start Building on AWS Today. Amazon Web Services. Available from: <https://aws.amazon.com/> (retrieved 4 June 2018).
- Amazon. 2017b. Amazon S3. Amazon Web Services. Available from: <https://aws.amazon.com/s3/> (retrieved 3 June 2018).
- Amazon. 2017c. Amazon EC2. Amazon Web Services. Available from: <https://aws.amazon.com/ec2/> (retrieved 4 June 2018).

- Baird, A., Jørgensen, S.H., Parada-Cabeliero, E., Hantke, S., Cummins, N., Schuller, B. (2017) Perception of Paralinguistic Traits in Synthesized Voices. In: Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences (AM 2017), London, UK, 23-25 August 2017. New York, USA: ACM. No pagination.
- Baird, A., Jørgensen, S.H., Parada-Cabeliero, E., Hantke, S., Cummins, N., Schuller, B. (2018) The Perception of Vocal Traits in Synthesized Voices: Age, Gender, and Human-Likeness. *The Journal of Audio Engineering Society*, Special Issue on Augmented and Participatory Sound and Music Interaction using Semantic Audio, 66 (4), 277-285.
- Barra-Chicote, R., Yamagishi, J., King, S., Montero, J. M., Macias-Guarasa, J. (2010) Analysis of statistical parametric and unit selection speech synthesis systems applied to emotional speech. *Speech Communication*, 52 (5), 394–404.
- Butler, J. (1993) *Bodies that Matter*. New York: Routledge.
- Clifford, N., L. Kwan Min. (2000) Does Computer-generated Speech Manifest Personality? An Experimental Test of Similarity-attraction. In: *Proceeding of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. New York, USA: ACM. 329–336.
- Mozilla Corporation. 2017. Project Common Voice. Mozilla Corporation. Available from: <https://voice.mozilla.org/da> (retrieved 4 June 2018).
- Dutoit, T., Leich, H. (1993) MBR-PSOLA: Text-To-Speech synthesis based on an MBE re-synthesis of the segments database. *Speech Communication* 13, 3–4.
- Feinberg, S., Murphy M. (2000) Applying Cognitive Load Theory to the Design of Web-based Instruction. In: *Proceeding of the 18th Annual ACM International Conference on Computer Documentation: Technology & Teamwork (SIGDOC '00)*. Cambridge, USA: IEEE. 353–360.
- FestVox. 2014a. Building Synthetic Voices: Building Prosodic Models. FestVox. Available from: <http://festvox.org/bsv/c1639.html> (retrieved 4 June 2018).
- FestVox. 2014b. Building Synthetic Voices: Corpus Development. FestVox. Available from: <http://festvox.org/bsv/c2176.html> (retrieved 4 June 2018).
- FestVox. 2014c. CMU Arctic. FestVox. Available from: [http://festvox.org/cmu\\_arctic/cmuarctic.data](http://festvox.org/cmu_arctic/cmuarctic.data) (retrieved 4 June 2018).
- Gagnon, R. (1978) Votrax real time hardware for phoneme synthesis of speech. In: *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'78)*. Tulsa, Oklahoma, USA. Cambridge, USA: IEEE, no pagination.
- Hall, S. (1994) Cultural Identity and Diaspora. In: Patrick Williams & Laura Chrisman (eds.). *Cultural Identity and Diaspora*. In *Colonial Discourse and Post-Colonial Theory: A reader*. New York: Columbia University Press, 392-404.
- Hunt, A.J., Black, A.W. (1996) Unit selection in a concatenative speech synthesis system using a large speech database. In: *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'96)*, Atlanta, GA, USA. Cambridge, USA: IEEE. 1520–6149.
- Huybrechts L. (ed.) (2014) *Participation Is Risky: Approaches to Joint Creative Processes*. Amsterdam: Valiz.
- Iida, A., Campbell, N., Higuchi, F., Yasumura, M. (2003) A corpus-based speech synthesis system with emotion. In: *Speech Communication*, 40 (1-2), 161–187.
- Jones, A. (2012) *Seeing Differently: A history and theory of identification and the visual arts*. New York: Routledge.
- Phan, T. (2017) The Materiality of the Digital and the Gendered Voice of Siri. *Transformations* 29, 23–33.
- Saratxaga, I., Navas, E., Hernáez, I., Luengo I. (2006) Designing and recording an emotional speech database for corpus based synthesis in Basque. In: *Proceedings of 5th international conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy. Luxembourg: European Language Resources Association (ELRA). 2126–2129.
- Schroeder, M. R. (1966) Vocoder: Analysis and synthesis of speech. *The Bell System Technical Journal*, 54 (5), 720–734.
- Peterson, G. E., Wang, W., Sivertsen, E. (1958) Segmentation techniques in speech synthesis. *The Journal of the Acoustical Society of America*, 30, 739–742.
- Tabet, Y., Boughazi, M. (2011) Speech synthesis techniques. A survey. *Proceedings of 7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA'11)*, Tipaza, Algeria. Cambridge, USA: IEEE. No pagination.
- Ward, W. (1989) Understanding Spontaneous Speech. In: *Proceeding of the Workshop on Speech and Natural Language (HLT '89)*.

- Stroudsburg, PA, USA: Association for Computational Linguistics. 137–141.
- Young, S. J. (1979) Speech synthesis from concept: A method for speech output from information systems. *The Journal of the Acoustical Society of America*, 66 (3), 685.
- Zen, H., Senior, A., Schuster, M. (2013) Statistical parametric speech synthesis using deep neural networks. In: *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'13)*. Vancouver, BC, Canada. Cambridge, USA: IEEE. No pagination.
- Oord, A. V. D., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K., (2016) Wavenet: A generative model for raw audio. arXiv, 1609:03499, Google DeepMind, London, UK.



La voix est dans l'air. Pas comme parole, pas comme chant,  
comme *voix* - qu'elle soit parlée, criée, fredonnée au  
chuchotée.

(Michel Chion, *La voix au cinéma*)

**Hear how your voices resonate,  
dissonate, appear, fluctuate,  
oscillate, interrelate, equivocate,  
overpower, overlap**





## Conclusion

As I have presented and discussed in the articles and texts included in this dissertation, vocal expressions found in art and digital media can create listening experiences and imaginaries that manifest, perform, negotiate and complicate auditive vocal bodies through technological manipulation, mediation and synthesis of paralinguistic expressions – but how do these technological possibilities actually enable performed vocal bodies? How do they manifest vocal expressions cementing paralinguistic-bodies within normative systems of e.g. male and female? And how do vocal expressions produce paralinguistic-bodies that open up for paralinguistic multivocalities?

In connection to the formulated common framework of my PhD project, I am specifically interested in how the relationship between technology and vocal paralinguistic expressions in art and digital media has stressed, complicated or altered experiences of paralinguistic-bodies and identities in relation to the listener. As noted, the paralinguistic-body can capture the many different relations between paralinguistic performance and notions of bodies and identities – including the imaginary of bodies, which to me presents a possibility of moving beyond the binary dichotomies of present or absent, embodied or disembodied, human or non-human voices.

I discuss the paralinguistic-bodies as they are unfolded as material, technological performative, visual and auditive in artworks and media performances in relation to the listeners experience and imaginaries. Although I am positioning myself as a listener, my listening is, as noted, not exclusively concerned with sound, but also with what is seen, moved, as well as imagined. As my listening and writing presented in the texts included in the combined PhD dissertation are situated and offer partial perspectives, my conclusion should therefore not be seen as providing any definitive answers, but rather be understood as a suggestive and open-ended engagement. In the following, I will discuss my previous interpretations of the vocal performances in digital media and contemporary art as can be read in the articles, documentation and remediation in relation to the concept of the paralinguistic-body. I will map out the different ways in which the vocal performances

researched in the works and projects included in this PhD project perform a range of vocal bodies from paralinguistic stereotypes to paralinguistic multivocalities.

### **Paralinguistic Stereotypes**

In the digital media performance, *THE END*, the audio-visual staging of the singing synthesis Hatsune Miku – which can be said to belong to the realm of computational paralinguistics – promotes an experience of a paralinguistic-body connected to a socially intelligible stereotype female character. This is stressed through the pairing of Hatsune Miku's little-girls-voice and the visual depiction of a 3D female bodied character with long legs in a mini skirt designed by Louis Vuitton. Hatsune Miku as a stereotypical female character is further stressed in the staging of Hatsune Miku as a counterpart to the male composer and DJ, Shibuya, who stands on stage behind a veiled mixer or electric piano seemingly orchestrating the vocalic appearance and paralinguistic expressions of the vocaloid. This citational practice of the gender binary can be understood as a way to make invisible the technological mediation of voice, making Hatsune Miku's paralinguistic performance appear live and present to listeners as a stereotype female paralinguistic-body. However, as I have argued, this also inscribes the performance of the vocaloid into structures of fetishism as a gender stereotype with a singular fixed (and sexualized) paralinguistic-body fully controlled and sculpted into the desire of the male artist on stage. Hatsune Miku, in this way, become a paralinguistic-fetish-body in binary subject-object relation, an object of desire for the empowered (male) subject's gaze and listening.

The staging of the Nao robot in the digital media performance *ROBOT* presents an alternative strategy of performing fixed paralinguistic-bodies. A Strategy that allows for different experiences of stereotypical representations of identities in comparison to the staging of Hatsune Miku's fetishized paralinguistic-body in *THE END*. This is especially clear in a scene where the little Nao robot with a 'boyish' appearance does calisthenics and flirts with a female dancer, asking "Don't you think I'm strong?" in a synthetic vocal tone, which also somewhat resembles that of a young boy. At first this appears to be a very stereotypical performance. However, the Nao robot's small, clumsy plastic body and slightly robotic voice do not enable the experience of the robot as 'someone' in the sense that LaBelle claims vocal paralinguistic performance does. Furthermore, the presumed attempt at performing the paralinguistic-body of a male stereotype within a human sociality

fails. Instead, the performance of this paralinguistic-body becomes a caricature; the calisthenics and vocal imitation of a male stereotype seems exaggerated. The reference to the male stereotype of a strong figure produces a grotesque and alienating effect as the robot-body is too small and fragile and the robot's vocal tone sounds too young to create a convincing performance of this stereotypical identity (which becomes especially clear in comparison to the female dancer who seems much stronger and bigger than the Nao robot). In this way, the clichéd scene creates a subversive experience where the normative stereotypes of male and female, human and non-human are questioned rather than reiterated. Just as LaBelle suggests the performative attempt of identity in vocal parody could do. The performance of Nao's caricatured paralinguistic-body is further complicated as Nao's robot-body sings with another voice, a stereotype female voice, later in the performance. The Nao robot's performances challenge the idea of the paralinguistic-body as a fixed identificatory body (as a 'grain' of voice operating within the bodily borders) performing two different stereotypes of paralinguistic-bodies. The subversive performance of both male and female gendered stereotypes connected to the Nao robot's one paralinguistic-body, highlights the absurdity of the gender performativity in humanoid robot technology and synthesized voice design.

### **Paralinguistic multivocalities**

Apart from the two digital media performances, THE END and ROBOT, operating with stereotyped paralinguistic-bodies, the rest of the works presented in this PhD project perform paralinguistic multivocalities. Here the 'body in the voice' transgresses the bodily borders (of Barthes' 'grain' of voice) in continual encounters with others (humans and nonhumans), as physical, imaginary, interrelational and becoming-with paralinguistic-bodies. The concept of paralinguistic multivocalities has been developed through my research into how technological paralinguistic-bodies can alter and complicate experiences of identificatory bodies in relation to listeners' experiences and imaginaries. Paralinguistic multivocalities should here be understood as heterogeneous forms of relationality that transgresses the bodily borders of vocal performance as something that is never fixed and confined to a singular identity or subjecthood. Paralinguistic multivocalities are something that is always divided, transgressed and becoming in relation

to other human and non-humans, which can be heard in the different works presented below.

In Pugholm's work *Emigration* (2013), paralinguistic expressions create a listening experience of an acousmatic paralinguistic-body, in which the body is separated from the voice heard: only clouds drift by on a blue sky in the video accompanying the vocal narration of an emigration story. Young writes that, in the acousmatic listening mode, the voice is still a strong metaphor for subjectivity but "no longer rooted in a particular time not to a specific locale" (2015, p. 27). The paralinguistic expressions are first performed as a paralinguistic-body connected to a stereotype Danish immigrant identity. Then the listener's experience of this paralinguistic-body is challenged by the interplay with the narrative told, as it is slowly revealed that the vocal expressions should not only be experienced as a representation the immigrants in Denmark, but also represent, just as much as, Danish emigrants. Through this double paralinguistic-body Pugholm's work questions the preconceptions of those identities apparently connected to an accented voice. The work plays with the listener's expectation and understanding of who the emigration story heard to belongs to, as this double paralinguistic-body moves the experience of the accented paralinguistic expression from being a marker of a specific and fixed identity to a sign of a transnational and transhistorical phenomenon – emigration. The acousmatic listening experience of the paralinguistic-body in *Emigrant* presents paralinguistics that are simultaneously situated and non-situated within a specific time and locality and it is in this way a performance of paralinguistic multivocality.

In Wagechi Mutu's *The End of Eating Everything* (2013), a paralinguistic-body unfolds through the performance of a monstrous scream that challenges the listening experience of the voice as merely coming from a human subject. In the video-work, the visual imagery shows that the scream is coming from a Medusa figure – half woman, half earth with eel-like hair curling up and stretching out to the dynamics of the scream. The monstrous scream creates the chilling ambivalence of this monstrous hybrid creature as an uncertain other – both human and non-human – as an experience of an unplaceable monstrous otherness which has a generative potential. The paralinguistic expression of the monstrous otherness is generative because the monstrous is other-than-human, but always in relation to the human in a way that reflects back on our humanity – which can make the listener rethink the human and "the production of otherness in the very *human*

being” (Kunst 2008, p. 2015) . The time-stretched scream is an expression connected to both the human and the non-human creating in the work a listening experience of a double paralinguistic-body. The scream as monstrous vocal otherness does not enable a listening experience of the intense paralinguistic expression as an identity as such. Rather it creates a listening experience of that which has been othered – this includes both humans, non-humans, the climate and the earth in the age of the Anthropocene – a paralinguistic multivocality. Through the scream listeners become engaged in a shared concern connected to consumption and mass production in the capitalist societies.

In Nanna Lysholt Hansen’s video work *Miss no. 1,2,3* (2013) a multi-voiced recitation by three identical clones are heard. The clones all look alike with long blond hair and white faces. Their voices are technologically manipulated and sound neither male or female, but as something that is hard to place within a gender binary system. The body and paralinguistic expressions are one and threefold simultaneously. The listener does not have the opportunity to define each clone as a subjectivity with its own identity, because the clones are pitched, reproduced and mediated through technologies that blend them together so that no single identity can be distinguished. The work challenges the auditive materiality of voices as markers of stereotypical gendered identities. Neumark writes that the voice is paradoxical because it embodies a contradiction, as both linked and detached from the body, both personal and interpersonal, as something that is neither subject nor object, and something that lies between the listener and the utterance (2010). LaBelle writes that the voice transgresses the bodily borders whereby classical subjectivity is defined, and instead lies in the in-between as fragmented, divided and partial (LaBelle 2014). The paralinguistic performances of the clones can be heard as deterritorializations: “turning one into *another*, as well as to hold oneself” (ibid., p. 98). The clone characters’ paralinguistic expressions do not create a listening experience of their paralinguistic-bodies as belonging to stereotypical binary identities, but instead create an experience of a paradoxical paralinguistic multivocality; of fragmented, divided, fluent technologically mediated identities ‘turning one into *another*’ as well as differentiating one from the other. In the interplay between the clones’ paralinguistic-bodies and the lists of how the word ‘miss’ is defined and categorized, an important problem of identity stereotypes is pointed out, namely that of categorization and naming. Through the technological stressing of the clone’s paralinguistic expressions as something that neither belong to male or female and

therefore cannot be placed within a binary system. The clones' vocal expressions point to how audio-visual bodies are alienated by the politics of categories and the ways language classifies and creates context for how one can identify and is identified and defined by others within a binary system, even when sounding and identifying with something that lies in-between or beyond those categories. Through the multivocal paralinguistic performances of the clones, this work stresses the political issue connected to the politics of naming and categorizing, and creates paralinguistic-bodies belonging to the realm structural critique rather than identity politics.

In another work by Lysholt Hansen, *Dear Daughter/Anatomy of the Chthulucene #6* (2018), an ambiguous connection between Lysholt Hansen's wireless loudspeaker and mouth (as a kind of double mouth) questions the presence of her paralinguistic expressions and makes the listener unsure from whom and where the voice heard emanates. The mimetic relationship between Lysholt Hansen's live performing body and its loudspeaker double are at the same time complementary and deconstructive. In the playback performance, it is sometimes difficult to detect what is live speech and what is pre-recorded speech. The loose synchronization and technologies of dubbing are destabilizing the paralinguistic-body as connected to a single fixed identity. It instead comes to exhibit partial identities that overlaps, merges, diffract, blends, etc. In this way a polyphonic entanglement of sonic kinship unfolds in Lysholt Hansen's vocal recitation. Different temporalities are merged – the past, the present, and the yet to come overlaps, synchronizes and diffuse through the performance of technologically mediated speech and live speech during the performance time. The technological paralinguistic-body is stretching Lysholt Hansen's vocal performance beyond the imagery of the singular body, manifesting many bodies of kinship. The multi-species kinship is manifested as a paralinguistic multivocality between mothers and daughters, female scientists, feminist thinkers and sexual tentacular organs. Through the multivocal paralinguistics Lysholt Hansen's performance gives voice to the silenced history of female sexuality and creates a space where female sexuality and intimate caring is fostered through multi-species mothering and making kin.

Another work in which multivocal paralinguistics connects to performances of paralinguistic-bodies of kinship is the video/shadow play *Transgressions* (2001) by the artist Nalini Malani. In the work, a dialogue of what could be between a mother and her

child, enables a listening experience of paralinguistic expressions that does not belong to a specific person but to many people. What is heard is the rhythm and tone of voice can be said to belong to what Neumark calls *enculturated bodies* (2010). The enculturated bodies connect to embodied voices that “are always already mediated by culture – modified by gender, ethnicity, and history” (Neumark 2010, p. 97) – but that are not pointing to one singular, stereotyped or fixed identity. The paralinguistic performance of the child-like voice in the work in fact questions the notion of fixed vocal identity; the age of the paralinguistic-body is hard to place as it is technologically modified; intense, high-frequency-sounding, age-ambiguous in its appearance to the extent that it poses questions as to whether the paralinguistic expressions heard belongs to the body of child at all. The accented paralinguistic performance is not located as belonging to the body of a clear-cut child or clear-cut not-child, it sounds in-between hard to locate as a singular identity, but still an enculturated paralinguistic-body. The technologically high-pitched paralinguistic-body interpolates the listener in a sonic experience stressing issues of globalization and consumerism and the consequences that these developments have in connection to enculturated others. In this way, the enculturated paralinguistic-body shows a multivocality that is not connected to identity politics, but rather – in an interplay with the visuals in the work showing different local Indian languages sinking in the ground – points to the structural problems connected to the ways in which the systems of globalization and consumerism enables or disables the voices of Others outside the dominant currencies and languages.

In the narrative experimental film, *The Woman, The Orphan and The Tiger* (2010) by Jane Jin Kaisen and Guston Sodin-Kung, a paralinguistic multivocality is also heard. A choir of three generations of women who have otherwise been silenced and traumatized in South Korean history is heard. The choir is interwoven so only fragments of words from each voice are heard. It is not the story of a single woman we hear, but the collective testimony of the choir; the experiences, conditions and emotions that are rooted in the body; the silencing, the oppression and the alienation. Each paralinguistic expression is a testimony of that which permeates the entire presence, but which cannot be captured in words. The technological layering of voices opens up an interrelational space, a plurality, through which the uniqueness of each vocal paralinguistic expression is communicated in a polyphony of other singular paralinguistic expressions. The choir penetrates the listener

with their urgent voices – and make the listener attend to the affective tonalities, rhythms, and intensities of the paralinguistic materiality of voices. The choir's affective paralinguistic expressions-in-relation becomes a testimony on a transnational and transhistorical phenomena – the traumatizing and silencing of these voices' stories, and their desire to be heard and inscribed in the national narratives. Connor writes that in the choralised voice "many-in-one becomes a kind of one-from-many" (2015, n.p.). The choir is a choral paralinguistic-body where the many vocal affective expressions create a body, a body of trauma, only made audible by through the many voices. The listener witnesses how the affective paralinguistic vocality express trauma, something which cannot justly be described in words. The choir performs a paralinguistic multivocality of narratives across the individual stories into a shared structural critique of transnational conditions and transhistorical issues in relation to the listener.

The collective vocal project *Collective Performative Reading* allows for an affective listening to and through paralinguistic expressions. As in *The Woman*, *The Orphan* and *The Tiger*, the affective choral paralinguistic-body are heard. Here, however, it takes shape as an explorative exchange between many-bodies-in-many-voices as a multivocality of dissonance and resonance between the participants present. The choral paralinguistic-body heard in the *Collective Performative Reading* is performed as expressions-in-relation through an inter-relationality. This interrelationality is not an ontological being in the world, but requires action and response. In *Collective Performative Reading*, there is a participatory element, where the people present are both listeners and creators of a momentarily shared paralinguistic expressions. The plural choral paralinguistic-bodies manifest through expressions-in-relation interconnecting the participants' identities constituting the chorality in equivocation. This 'acting in concert' also resonates in the *Collective Performative Reading*, as coordinated action and 'orchestrated collectivity' of voices 'exercising a plural and performative right to appear' (Butler, 2015). The *Collective Performative Reading* challenge the listening experience of binary-oppositional notions of identity as singular through the choral paralinguistic-body as inter-relationality, equivocation and multivocality.

The collective vocal project [*multi'vocal*] creates a paralinguistic multivocality as an investigation into the alternatives to conventional stereotypes found in synthetic voice designs, exemplified here in relation to vocaloid Hatsune Miku and the Nao robot. In the



[*multi'vocal*], synthetic voice there are many-bodies-in-one-voice which can be heard as a choral paralinguistic-body. Cavarero's conceptualization of the voice as something that has an auditive meaning on its own, is prevalent in the [*multi'vocal*] voice, which does not really perform meaningful semantics, but instead creates meaning through the paralinguistic multivocality heard. This is done by displaying an affective vocal tone embodying many genders, ages and accents recorded from different participants, as an alternative to the singular stereotyped paralinguistics of synthetic voices. Connor explains how chorality can go beyond itself in amplitude as volume: "the choral voice is the voice that goes beyond itself [...] It is not just the body's power of emanation, it is emanation's power to grow into a kind of hyper-body" (2015, n.p.). In [*multi'vocal*], the synthetic voice is not amplitude understood as volume, but amplitude understood as dimensions that turn the choralic paralinguistic-body into a hyper-body. The [*multi'vocal*] synthetic voice is a hyper-body, a proliferation of vocal paralinguistic expressions and an exploration of the possibilities of listening beyond dualisms such as male and female. It is an exploration of how digital vocal design can open up for the trading duality for a multiplicity that, as the queer feminist scholar, Mimi Marinucci, writes in her book, *Feminism is Queer: The Intimate Connection Between Queer and Feminist Theory* (2016), queer identifications affords. The [*multi'vocal*] synthetic voice is a choric voice in that give rise to "the fantasy of a collective voice-body that is not to be identified with any of the individuals who compose it [...] It is the sonorous actualising of the otherwise abstract or merely attributive idea of a collectivity" (Connor, 2015, n.p.). The [*multi'vocal*] synthetic voice is a choric paralinguistic-hyper-body creating an imaginary of a collective multivocal paralinguistic-body and is in this way a design with an alternative 'quantification of vocal tone' to that of the synthetic paralinguistic stereotypes such as those exemplified by Hatsune Miku or Nao's voice.

## Summing Up and Pointing Forward

In the globalized, networked, diasporic world of the twenty-first century, the dichotomy of the oppositional structure of the self/other simply doesn't hold explanatory value any more – particularly for younger generations who live on screens or mobile phones and who routinely cross boundaries of class,

race, gender, sexuality, and so on in their personal relationships and performative processes of self-defining (Jones, 2012, p. 222).

The voices of commercial technologies like e.g. Hatsune Miku and Nao seem to be designed within well-known established dichotomies. However, as art historian Amelia Jones argue in the quote above, more and more people do not identify with binary notions of identity and so there is a need to find, as already noted, alternative ways of designing and performing e.g. paralinguistic expressions of synthetic voices beyond a binary mode of representation.

Through my focus on paralinguistic technologies in this PhD dissertation, I have highlighted the commonalities of the different explorations presented in the articles and texts. I have done so by showing how they all contribute to a shared field of study, generating a new relevant perspective on the discussions of the auditive, technological and performative aspects of paralinguistics within sound studies and computational paralinguistics. Building on the three articles and the collective vocal projects included in the dissertation, I have explored different ways in which paralinguistic expression perform vocal auditive bodies in art and digital media in relation to listeners. In my investigations, I have listened to and through vocal paralinguistic performances. Through this I have, like LaBelle, mapped different paralinguistic performances of vocal materiality “in and around the spoken” (2014, p. 183).

In this dissertation, I have developed the term paralinguistic-body and then presented, discussed and mapped how the performance of different technological paralinguistic-bodies stereotype, alter and expand vocal bodies. First, I discussed paralinguistic expressions of synthetic voices that I argued create listening experiences of stereotyped paralinguistic-bodies with the aim to make technological mediation unheard – staging (and subverting) the synthetic paralinguistic expressions within a gender binary dichotomy. Then I discussed how technologically mediated and manipulated paralinguistic expressions can present paralinguistic-bodies that challenge paralinguistics as marker of a fixed vocal body (within the framework of computational paralinguistics – as a biological trait primitive) performing multivocal paralinguistic-bodies.

I have been invested in listening to the performance of paralinguistic-bodies as material performances of vocal bodies as physical, technological, imagined, human and

non-human – expanding the boundaries of what is considered a vocal body. Not as how Barthes' 'grain' of voice operates within the bodily borders of a human subject, but rather as a 'paralinguistic-body' combining Connor's 'voice-body' – manifested as physical and imaginary projection of voice and body – and LaBelle's performative 'body in the voice' where the paralinguistics are relational and continuously defined in the encounter with the other (developing Cavarero's theory of vocal auditive difference). I use the paralinguistic-body as a term to direct attention to how the material auditive performances create experiences and imaginaries of bodies, not as singularly fixed, but as active interrelationalities between vocal productions and listeners.

The development of the term paralinguistic-body can contribute to the field of computational paralinguistics by focusing on the dynamic interrelationality of paralinguistic expressions and listeners. In the articles included, I discuss more or less explicitly how vocal expressions are mediated, modified and manipulated through technologies and how this manifest, transform, alter and complicate experiences of paralinguistic-bodies in relation to the listener. The discussion on the paralinguistic multivocality contributes to the field of paralinguistics unfolding how paralinguistic states are stressed and transformed through technological mediation, distribution or synthesis. Performing vocal bodies that are in-relation, split, partial, becoming-with and connected to concerns and issues move listening experiences of vocal sonic performance away from identity politics to interrelationalities, structural critiques or imaginaries of alternative worlds and conditions.

My focus on the technological as something that alters and stresses the experiences of paralinguistic expressions of auditive bodies elaborates on LaBelle's 'cultures of the paralinguistic' to include the technological and computational aspects of paralinguistics. This expands the maps the connection to experiences of technologically modified, sculpted, mediated, synthesized, transformed and stressed paralinguistic expressions. I also develop the definition of traits and states within computational paralinguistics by including my listening experience of paralinguistic expressions in the analysis. I have shown how my listening experience, imagination, associations, interpretations and understanding of technological paralinguistic performances cannot be taken out of the equation when evaluating paralinguistic qualities and how a sound studies perspective aiming to listening beyond dualisms can contribute to the analysis of paralinguistic

expressions in computational paralinguistics (including the listeners' experiences and understandings of the paralinguistic expressions). By linking sound studies with computational paralinguistics, this PhD project is part of a consolidation of the interdisciplinary field of paralinguistic technologies.

Within the interdisciplinary field of paralinguistic technologies, this PhD project also show new collective ways of performing and designing paralinguistics in art and digital media as it is explored in the *Collective Performative Reading* and the *[multi'vocal]* project. In *[multi'vocal]*, the possibility is investigated of creating a participatory, fragmented, ever-evolving and pluralistic synthesized voice trained on thousands of voice recordings from different participants, as an alternative to the commercial normative designs of synthesized vocal paralinguistic stereotypes (as it can be experienced with the staging of e.g. Hatsune Miku and Nao). A question that comes forth through *[multi'vocal]* is whether paralinguistic tone of voice connected to gender, age and accent should be considered as long term biological trait primitives or as paralinguistic short term states, "as a mode and speaking style" in computational paralinguistic (Schuller et al, 2013, p. 7). The evaluation techniques and categorizations of paralinguistic expressions as states or traits is also an important future area of study. In the *[multi'vocal]* collective we have made a preliminary study into how to evaluate the *[multi'vocal]* synthetic voice different from the perception tests grounded in statistical research with a qualitative approach conducting semi-structured interviews. Here vocal sonic qualities were discussed in relation the listeners' imaginaries as part of the perception test. Apart from defining the computational paralinguistic expressions as male and female sounding, informants heard "the sound of the ocean and the sea" and "water roaring" (Juutilainen, 2019, p. 52) as well as "mechanical" sounds connected to "robots" (ibid., p. 53). The paralinguistic-body of the *[multi'vocal]* synthetic voice performed a paralinguistic multivocality as organic and mechanical, male and female to informants. Further, the listeners' descriptions of the same computational paralinguistic expressions could change throughout the sample heard, and suggests that what is considered traits in traditional computational paralinguistics could possibly be experienced and defined as temporary states. This preliminary exploration of how to evaluate listener experiences of digital paralinguistic expressions including the listener associations and

situatedness, points to other ways conducting and reflecting upon listener experiences of e.g. synthetic voices as an area for future research.

The political, economic and social aspects of paralinguistic technologies are also all important future areas of research. In relation to the research presented here, the socio-technical aspect of paralinguistic technologies are the next steps. Questions as to whether synthetic voices can be experienced as having more than one voice are still to be explored further through design and listening to and through new alterations of the *[multi'vocal]* synthetic voice as well as other participatory, collective or community-based synthesized voices. These are the important areas for future research within paralinguistic technologies.



Its a fine day • dust crowds pay my mouth a visit • dust  
voices like mine yet not mine • ringing in the hol • dust  
voices ash clouds in my mouth on this fine day • ðes fine  
day ðes fine dæg • ash clouds of the dead the dust bein of  
the daudr on ðes fin dæg • ðes fine day • the dust bân of  
the daudr • fine day fin dæg • fill my mouth up • ring in  
min muð on ðes fin dæg • open my muð for the þing of the  
hol • the þing of my mouþ for the sound of the hol • listen  
here hear the sound on ðes dæg • ring on ðes day • nu  
listen nu here hear ring on ðes fin dæg • denne fine dæg

(Caroline Bergwall: *Drift*)

**You may keep the quote, pass it on to another or put it back on the floor or somewhere else for someone else to possibly read and perform together**





## References

- Austin, J. L. (1962) *How to Do Things with Words*. Oxford: Clarendon Press.
- Baird, A., Jørgensen, S.H., Parada-Cabaleiro, E., Cummins, N., Hantke, S., Schuller, B. (2017) Perception of Paralinguistic Traits in Synthesized Voices. *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*, pp. n.p.
- Baird, A., Jørgensen, S.H., Parada-Cabaleiro, E., Cummins, N., Hantke, S., Schuller, B. (2018) The Perception of Vocal Traits in Synthesized Voices: Age, Gender, and Human-Likeness. *JAES Journal of the Audio Engineering Society*, Vol. 66 (4), pp. 277-285.
- Barthes, R. (1977) The Grain of the Voice. *Image-Music-Text*. Translated by S. Heath. London: Fontana Press, pp. 179-189.
- Barthes, R. (1985) Listening. *The Responsibility of Forms: Critical Essays on Music, Art, and Representation*. Translated by R. Howard. New York: Hill and Wang, pp. 245-260.
- Bickmore, T.W., Cassell, J. (2005) Social Dialogue with Embodied Conversational Agents. In: J. v., Kuppevel, L. Dybkjaer, and N. Bernsen, eds. *Advances in Natural Multimodal Dialogue Systems*, Dordrecht: Kluwer Academic, pp. 23-54.
- Boss, S., 2018. *Tuning the Ear - Exploring Conditions and Conceptions of Hearing*. PhD Dissertation, Faculty of Arts, Aarhus University.
- Brautigam, B. (2017) The New Skeuomorphism is in Your Voice Assistant. [Online] Available at: <https://uxdesign.cc/the-new-skeuomorphism-is-in-your-voice-assistant-3b14a6553a0e> [Accessed: November 16th 2019].
- Butler, J., Athanasiou, A. (2013) *Dispossession: The Performative in the Political*. Malden: Polity Press.
- Butler, J. (2015) *Notes Toward a Performative Theory of Assembly*. Cambridge: Harvard University Press.
- Busch, K. (2009) Artistic Research and the Poetics of Knowledge. *Art & Research*, Vol. 2 (2) pp. 1-7.
- Buzzanell, P. M. (2017) Voice/Multivocality. In: J. Matthes, C.S. Davis and R. F. Potter, eds. *The International Encyclopedia of Communication Research Methods*. Hoboken: John Wiley & Sons.

- Candy, L. (2006) Practice Based Research: A Guide. *CCS Report*. [Online]. Available at: <http://www.creativityandcognition.com/resources/PBR%20Guide-1.1-2006.pdf> [Accessed: November 16th 2019].
- Candy, L., Edmonds, E. (2018) Practice-Based Research in the Creative Arts. *Leonardo*, 51(1) pp. 63-69.
- Cavarero, A. (2005) *For More Than One Voice: Toward a Philosophy of Vocal Expression*. Stanford: Stanford University Press.
- Cermak-Sassenrath, D., A. Okutsu, S. Hasse. (2015) Electromagnetic Landscape: In-between signal, noise and environment. *ISEA 2015: Disruption Proceedings*, pp. n.p.
- Cermak-Sassenrath, D., Jørgensen, S. H. (2018) Re-Volution Sampler - A Participatory Archive of Revolutionary Songs. *The 24th International Symposium on Electronic Art (ISEA) 2018: Intersections Wrap-Up*, p. 59.
- Charalambides, S. (2020) When the common ground is shattered: Self-enclosed individualism and partial relationality in creative practice. *Peripeti*, forthcoming.
- Chion, M. (1990) *Audio-Vision: Sound on Screen*. Translated by C. Gorbman. New York: Columbia University Press.
- Connor, S. (2000) *Dumbstruck: A Cultural History of Ventriloquism*. Oxford: Oxford University Press.
- Connor, S. (2014) *Beyond Words: Sobs, Hums, Stutters and other Vocalizations*. London: Reaktion Books.
- Connor, S. (2015) Choralities. [Online]. Available at: <http://stevenconnor.com/choralities.html> [Accessed: November 16th 2019].
- Crawford, K., Joler, V. (2018) Anatomy of an AI System: The Amazon Echo as an anatomical map of human labor, data and planetary resources. [Online]. Available at: <https://anatomyof.ai> [Accessed: November 16th 2019].
- Crystal, D. (1975) Paralinguistics. In: J. Benthall and T. Polhemus, eds. *The Body as a medium of expression: essays based on a course of lectures given at the Institute of Contemporary Arts*. London: Allen Lane, pp 162-174.
- Cvejić, B. (2005) Collectivity? You Mean Collaboration. *Republicart: Artists as Producers*. [Online]. Available at: [republicart.net/disc/aap/cvejic01\\_en.htm](http://republicart.net/disc/aap/cvejic01_en.htm) [Accessed: November 16th 2019].

- Dardot, P., Laval, C. (2013) *The New Way of the World: On Neoliberal Society*. New York: Verso.
- Derrida, J. (1967/1997) *Of Grammatology*. Translated by Gayatri Chakravorty Spivak. Baltimore: John Hopkins University Press.
- DR Nyheder (2018) [Online]. Available at: <https://www.dr.dk/nyheder/viden/teknologi/vild-google-demo-kunstig-intelligens-ringer-til-frisoeren-og-bestiller-tid> (May 8th, 2018). [Accessed: November 16th 2019].
- Dyson, F. (2014) *The Tone of Our Times: Sound, Sense, Economy, and Ecology*. Cambridge: MIT Press.
- For More than One Voice. (2020) Collective Performative Reading (remediation). *Peripeti*, forthcoming.
- Gagnon, R. T. (1978) Votrax real time hardware for phoneme synthesis of speech. *ICASSP 1978*, pp. 175-178.
- Haraway, D. (1988) Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, Vol. 14 (3), pp. 575-599.
- Harris, R. A. (2005) *Voice Interaction Design: Crafting the New Conversational Speech Systems*. San Francisco: Morgan Kaufmann Publishers /Elsevier.
- Hasse, S. (2012) I am (recording) sitting in a room. *Body, Space & Technology*, Vol. 11.
- Holmboe, R. (2019) *The Resonant Museum: Sound, Art and the Politics of Curating*. PhD Dissertation, Department of Arts and Cultural Studies, University of Copenhagen.
- Hunt, A., Black, A. (1996) Unit selection in a concatenative speech synthesis system using a large speech database. *Proceedings ICASSP*, pp. 373-376.
- Ihde, D. (1976/2007) *Listening and Voice: A Phenomenology of Sound*. Albany: State University of New York Press.
- Jalving, C. (2011) *Værk som Handling*. København: Museum Tusculanum.
- Jansen, L., Westphal, M. (2017) Rihanna Works Her Multivocal Pop Persona: A Morpho-syntactic and Accent Analysis of Rihanna's Singing Style: Pop culture provides rich data that demonstrate the complex interplay of World Englishes. *English Today*, Vol. 33 (2), pp. 46-55.
- Jones, A. (2012) *Seeing Differently: A History and Theory of Identification and the Visual Arts*. New York: Routledge.

- Juutilainen, F.T. (2019) *Multivocal - Creating synthetic voices with non-singular identities*. Thesis, Department of Nordic Studies and Linguistics, University of Copenhagen.
- Jørgensen, S. H., Snyder, J. (2012) Move/Bevæg Dig: An Interactive Sound Art Installation. *Music, Mind, and Invention Workshop*, pp. n.p.
- Jørgensen, S. H., Tafdrup, O. (2017) Technological Fantasies of Nao - Remarks about Alterity Relations. *Transformations*, 29, pp. 88-103.
- Jørgensen, S. H., Vitting-Seerup, S., Wallevik, K. (2017) Hatsune Miku: An uncertain image. *Digital Creativity*, Vol. 28(4), pp. 318-331.
- Jørgensen, S. H. (2018) ROBOT and THE END: A Comparative Critical Reading of the Staging of Synthesized Voices in Digital Media Performances. *Peripeti*, Vol. 15 (29/30), pp. 156-169.
- Jørgensen, S. H., Baird, A.E., Juutilainen, F.T., Pelt, M., Højholdt, N.C. (2018) [multi'vocal]: Reflections on Engaging Everyday People in the Development of a Collective Non-Binary Synthesized Voice. *Proceedings of EVA Copenhagen, British Computer Society*. pp. 1-8.
- Kahn, D. (1999) *Noise, Water, Meat: A History of Sound in the Arts*. Cambridge: MIT Press.
- Kelly, M., Widenheim, C. (2012) *Dialogue - On the Politics of Voice*. Stockholm: Moderna Museet.
- LaBelle, B. (2012) Acoustic Spatiality. *Broj 4 / Književnost i kultura*.
- LaBelle, B. (2014) *The Lexicon of The Mouth: Poetics and Politics of Voice and the Oral Imaginary*. London: Bloomsbury Academic.
- Lane, C. (2006) Voices from the Past: Compositional Approaches to using recorded speech. *Organised Sound*. Vol. 11 (1), pp. 3-11.
- Lawy, J. (2017) Theorizing voice: Performativity, politics and listening. *Anthropological Theory*. Vol. 17 (2), pp.192-215.
- Lee, E. J., Nass, C., Brave, S. (2000) Can Computer generated Speech Have Gender?: An Experimental Test of Gender Stereotype. *Conference on Human Factors in Computing Systems (CHI)*, pp. 289-290.
- Loxley, J. (2007) *Performativity*. New York: Routledge.
- Marinucci, M. (2016) *Feminism is Queer: The intimate connection between queer and feminist theory*. New York: Zed Books.

- Meinedo, H., Trancoso, I. (2010) Age and gender classification using fusion of acoustic and prosodic features. *Proceedings of Interspeech. ISCA*, pp. 2818–2821.
- Mizzi, R. (2010) Unraveling Researcher Subjectivity Through Multivocality in Autoethnography. *Journal of Research Practice*, Vol. 6 (1), pp. 1-13.
- Männistö-Funk, T., Sihvonen, T. (2018) Voices from the Uncanny Valley How Robots and Artificial Intelligences Talk Back to Us. *Digital Culture & Society*, Vol. 4 (1), pp. 45-64.
- Nancy, J-L. (2002/2007) *Listening*. Translated by Charlotte Mandel. New York: Fordham University Press.
- Nass, C., Brave, S. (2005) *Wired for Speech: How Voice Activated and Advances the Human- Computer Relationship*. Cambridge: MIT Press.
- Nass, C., Moon, Y., Green, N. (1997) Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers with Voices. *Journal of Applied Social Psychology*, Vol. 27 (10) pp. 864-876.
- Nass, C., Foehr, U., Brave, S., Somoza, M. (2001) The Effects of Emotion of Voice in Synthesized and Recorded Speech. *Association for the Advancement of Artificial Intelligence (AAAI) Technical Report*.
- Neumark, N. (2010) Doing Things with Voices: Performativity and Voice. In: N. Neumark, R. Gibson and T. v. Leeuwen, eds. *VØICE: Vocal Aesthetics in Digital Arts and Media*. Cambridge: MIT Press. pp. 95-118.
- Neumark, N. (2010) Introduction: The Paradox of Voice. In: N. Neumark, R. Gibson and T. v. Leeuwen, eds. *VØICE: Vocal Aesthetics in Digital Arts and Media*. Cambridge: MIT Press. pp. xv-xxxiii.
- Novak, J. (2015) *Post-Opera: Reinventing the Voice-Body*. Farnham: Ashgate Publishing.
- Collective/Kollektiv (2020) *Peripeti*, Forthcoming.
- Phan, T. (2017) The Materiality of the Digital and the Gendered Voice of Siri. *Transformations*, Vol. 29, pp. 23-33.
- Picard, R.W. (1997) *Affective computing*. Cambridge: MIT Press.
- Pinch, T., Bijsterveld, K. (2004) Introduction to Sound Studies: New Technologies and Music. *Social Studies of Science*, Vol. 34 (5), pp. 635-648.
- Repetto, D. (2010) Doing Things Wrong. *Frontiers of Engineering: Reports on Leading - Edge Engineering from the 2010 Symposium 2010*, pp. 47-50.

- Rogoff, I. (2002) WE – Mutualities, Collectivities, Participations. In: ed. *I Promise It's Political*. Cologne: Museum Ludwig, pp. 126-134.
- Schaefer, M. R. (1977) *The Tuning of the World*. New York: Knopf.
- Schuller, B., Batliner, A. (2014) *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. Hoboken: John Wiley & Sons.
- Schuller, B., Elkins, A., Scherer, K. (2017) Computational Analysis of Vocal Expression of Affect: Trends and Challenges. In: J. Burgoon, N. Magnenat-Thalmann, M. Pantic, and A. Vinciarelli, eds. *Social Signal Processing*. Cambridge: Cambridge University Press, pp. 56-68.
- Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., Narayanan, S. (2013) Paralinguistics in speech and language – state- of-the-art and the challenge. *Computer Speech & Language. Special Issue on Paralinguistics in Naturalistic Speech and Language*, Vol. 27 (1), pp. 4-39.
- Schulze, H. (2015) The Body of the Voice. *Seismograf*. [Online]. Available at: <https://seismograf.org/en/artikel/the-body-of-the-voice> [Accessed: November 16th 2019].
- Schulze, H. (2018) *The Sonic Persona: An Anthropology of Sound*. New York: Bloomsbury Academic.
- Schlichter, A. (2011) Do Voices Matter? Vocality, Materiality, Gender Performativity. *Body and Society*, Vol. 17 (1), pp. 31-52.
- Schmitt, M., Schuller, B. (2015) Machine-based decoding of paralinguistic vocal features. In: eds. S. Frühholz and P. Belin. *The Oxford Handbook of Voice Perception*. Oxford: Oxford University Press.
- Smith, H., Dean, R. (2009) *Practice-led research, research-led practice in the creative arts*. Edinburgh: Edinburgh University Press.
- Sterne, J., ed. (2012) *Sound Studies Reader*. New York: Routledge.
- Tamagawa, R., Watson, C., Han Kuo, I., MacDonald, B., Broadbent, E. (2011) The Effects of Synthesized Voice Accents on User Perceptions of Robots. *International Journal of Social Robotics*, Vol. 3 (3) pp. 253-262.
- Tompkins, D. (2011) *How to Wreck a Nice Beach: The Vocoder from World War II to Hip-Hop, The Machines Speak*. New York: Melville House Publishing.

- Trager, G. L. (1958) Paralanguage: a first approximation. *Studies in Linguistics*, Vol. 13, pp. 1-12.
- Verran, H. (1999) Staying True to the Laughter of Nigerian Classrooms. In: J. Law and J. Hassard, eds. *Actor Network Theory and After*. Oxford: Blackwell Publishers.
- Vallee, M. (2017) Technology, Embodiment, and Affect in Voice Sciences: The Voice is an Imaginary Organ. *Body and Society*, Vol. 23 (2), pp. 83-105.
- Watts, O., Henter, G. E., Merritt, T., Wu, Z., King, S. (2016) From HMMs to DNNs: where do the improvements come from? *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE*, pp. 5505-5509.
- Weidman, A. (2014) Anthropology and Voice. *Annual Review of Anthropology*, Vol. 43 (1), pp. 37-51.
- Wood, M. (1977) Deconstructing Derrida. [Online]. Available at: <https://www.nybooks.com/articles/1977/03/03/deconstructing-derrida/> [Accessed: November 16th 2019].
- Yen, C., Nass, C. (2010) *The Man Who Lied to His Laptop: What We Can Learn About Ourselves from Our Machines*. New York: Penguin Group.
- Yoffie, D.B., Wu, L., Sweitzer, J., Eden, D., and Ahuja, K. (2018) Voice War: Hey Google vs. Alexa vs. Siri. [Online]. Available at: <https://www.acetechbc.ca/wp-content/uploads/2019/03/Voice-War-Hey-Google-vs.-Alexa-vs.-Siri.pdf> [Accessed: December 16th 2019].
- Young, M. (2015) *Singing the Body Electric: The Human Voice and Sound Technology*. Farnham: Ashgate Publishing.
- Zen, H., Senior, A., Schuster, M. (2013) Statistical parametric speech synthesis using deep neural networks. *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on IEEE*, pp. 7962-7966.



## Abstract

Voice is both a matter of expression and of being heard, and connects deeply to feelings of intimacy, identity and bodily practices and imaginaries. With this PhD project *Vocal Bodies: Performing Paralinguistic Stereotypes and Multivocalities in Art and Digital Media*, I investigate how vocal expressions are stressed and transformed through technological mediation, distribution or synthesis. Through my methodological approach of listening to and through vocal expressions, I map out how technologies shape vocal stereotypes or multivocalities. Here I specifically focus on the paralinguistics, that is the auditive and bodily aspects of voice alongside linguistics. The discussion presented in the dissertation has been developed through an interplay between the theory and practice that has unfolded during the PhD project. Grounded in my analysis of different contemporary artworks and in the practice-led research conducted in relation to the collective vocal projects *Collective Performative Reading* and *[multi'vocal]*, I argue that paralinguistic performances can be multivocal, moving experiences of vocal sonic performance away from associations with identity politics to that of structural critiques or imaginaries of alternative worlds and conditions – in an interplay with the listener. I present several different ways in which paralinguistic multivocality is performed in art and digital media as auditive bodies of transnational and transhistorical trauma, emigration, consumption, gendered oppression, and othering in the neoliberal consumer society. Finally, the discussion presented in the PhD dissertation also present new collective ways of designing paralinguistics in digital media. The studies on technological paralinguistics conducted in this dissertation contribute to the field of sound studies, connecting it to audio engineering and computer science. This is an important new field of study as the paralinguistics of voices whether real or imaginary, are part of shaping how we come to understand ourselves and others through a material relationality of vocal expressions.

## Resumé

Stemmen er både et spørgsmål om udtryk og at blive hørt, og forbindes dybt med følelser af intimitet, identitet og kropslig praksis og forestillinger. Med dette ph.d.-projekt *Vokale Kroppe: Performance af Paralingvistiske Stereotyper og Multivokaliteter i Kunst og Digitale Medier* undersøger jeg, hvordan vokale udtryk understreges og transformeres gennem teknologisk mediering, distribution eller syntese. Diskussionen præsenteret i afhandlingen er udviklet gennem et samspil mellem teori og praksis, der er udfoldet i løbet af ph.d.-projektet. Min metodologiske tilgang om at lytte til og gennem vokale udtryk, bruger jeg til at kortlægge hvordan teknologier kan forme vokale stereotyper eller multivokaliteter. Her fokuserer jeg specifikt på paralingvistikken - de auditive og kropslige aspekter af stemmen. Baseret i min analyse af forskellige nutidige kunstværker og i den praksisledte forskning, der er foretaget i relation til de kollektive vokale projekter *Collective Performative Reading* og *[multi'vocal]*, argumenterer jeg for, at performativ paralingvistik kan være multivokal, som noget der bevæger oplevelser af performative vokale kroppe væk fra identitetspolitik og i stedet til strukturelle kritikker eller forestillinger om alternative verdener - i et samspil med lytteren. Jeg præsenterer her adskillige forskellige måder, hvorpå paralingvistisk multivocalitet udføres i kunst og digitale medier som vokale kroppe for tværnationale og transhistoriske traumer, emigration, forbrug, undertrykkelse af kønnene og andethed i det nyliberale forbrugersamfund. Endelig præsenterer diskussionen i ph.d.-afhandlingen også nye kollektive måder at designe vokale kroppe i digitale medier. Undersøgelserne om performativ teknologisk paralingvistik, der er foretaget i denne afhandling, bidrager til området sound studies og forbinder det med felter som lydteknik og datalogi. Det er et vigtigt nyt felt, da studiet i vokale kroppe og stemmens performative paralingvistik, hvad enten den er virkelig eller imaginær, er en del af den måde hvorpå vi lærer os selv og andre at kende.

# Appendix

## The politics of voices in contemporary art

Everywhere there are voices. Voices from others and from ourselves. Some pass by us, while others are concerning and intrusive. As voices speak, we feel the rhythms and sonic textures. Notions of intimacy, identity, authority and belonging are formed in the space between the voices and those who listen to them. Our inner voice mixes with the voices we hear.<sup>1</sup>

The voice is always interrelational. We cannot be above, below or outside the voices. When we listen to the speech of others, our experience of them, the way they speak, the accent they have, the words they choose to use, are all shaped by the surrounding social, cultural and political norms. Because of the importance and omnipresence of voices they have also become a subject and media for contemporary art. In this contribution, I will present selected Danish contemporary artworks and discuss how they, in different ways, work with the politics of voices in a dialogue with the listener. This contribution was made on the basis of a longer collaboration with Jane Jin Kaisen on the project *For More Than One Voice* (2016 -), which deals with the politics and poetics of the polyphony of voices, as well as through discussions and interviews with the artists Niels Pugholm og Nanna Lysholt Hansen.

### The voice as artistic media

The voice theorist Nori Neumark writes in the book *VØICE* from 2010 that today the voice has returned to the international theoretical and artistic agenda. Since the beginning of 2000, a host of books, articles and works of art that treat the voice as a linguistic, sonic and bodily phenomenon have been published and exhibited (Neumark 2010: xv).<sup>2</sup>

Sound theorists and artists like Mary Kelly, Miriama Young, Lawrence Abu Hamdan and Brandon LaBelle have written about the uniqueness of the voice as an artistic medium. Here is a focus on how the voice can create a dialogue between the work and the listener. In the book, *Dialogue - On the Politics of Voice*, Kelly writes that when we give attention to the voice in art, it is like concentrating sharply on the perspective of a painting or the movements and attire of a performance artist as essential elements in our experience of the creation of the work (Kelly 2011: 12). The voice similarly creates significance in the dialogue between work and viewer. Here, the viewer is also a listener who interprets the use of voice in a work, in the same way that a spectator registers the use of perspective and observes movements, colors and composition in a work and understands them as essential elements in the meaning-making of art. Young writes in the book *Singing the Body Electric* from 2015 that the voice can capture the listener with its evasive qualities: as a listener, one can be drawn by the material, sensuous expressions of the voice and reflect upon its metaphorical meanings. But we can never fully capture the voice, she writes, and it is herein that its artistic potential lies (Young 2015: 2). With the article *Aural Contract: Forensic Listening and the Reorganization of the Speaking Subject* from 2012, Hamdan also draws attention to the listener's role in relation to the voice. He wants to generate a discussion about how we listen to voices and how this listening is linked to broader political and social interests, which is also reflected in Hamdan's works such as *The Whole Truth* (2012) or *Conflicted Phonemes* (2012). These works

---

<sup>1</sup> Curator and art theorist, Irit Rogoff writes in the 2004 article, *We - Collectivities, Mutualities, Participations*, that the meanings an artwork produces are something that is created and shared collectively through the circulation of meaning and relationships that unfold around the artwork. This collectivity constitutes a 'we' in relation to the artwork, which is what is referred to by the 'we' that appears in this article.

<sup>2</sup> Books with the voice as a philosophical and theoretical focal point include Steven Connor's *Dumbstruck: A Cultural History of Ventriloquism* from 2000, Adriana Cavarero's *For More Than One Voice* and Mladen Dolar's *A Voice and Nothing More*, both from 2006.

problematize the emergence of voice analysis devices as investigative tools within judicial bodies. In *Conflicted Phonemes*, for example, Hamdan examines how particular ways of listening are associated with specific social and political understandings of the speaking subject. In the work, Hamdan explores, among other things, how voice analysis devices require forensic listening methods that categorize the speaking subject within a one-sided system of categories that cannot accommodate for the ambiguity and variability of the voice. In the work *Channels* (2013), created by the artist Susan Hiller, we hear stories of different people's experiences of death. Here, the voices are linked in an exploration of our cultural experiences of how we listen to voices as testimonies. The voice as a testimonial is explored in another way in Arthur Watson's conceptual performance *Singing for Dead Singers* (2000), in which Watson sings ballads on a Scottish dialect that exists only orally and which will be silenced unless the songs are sung as a manifestation of this dialect and musical tradition.

In the book *Background Noise: Perspectives on Sound Art* from 2006, LaBelle writes that the use of voice in art can often be linked to a political and social artistic practice. An art that, among other things, addresses issues of identity and belonging, which relate to how one is constituted and constructed as a subject or object, and how this is positioned and performed in the social and political spectrum (LaBelle 2006: 101). This can be argued to apply to Hamdans, Hillers and Watson's works, and it can also be said to apply within the Danish contemporary art scene. One can, for example, experience this in works such as *The Image of Me* (2012) by Jeanette Ehlers and Patricia Kaersenhout, Hannah Heilman's TV-art *Allons-y* (2016) and Kristoffer Ørum's *Data bodies* (2016). These works variously stage a listening of the voice as something that can destabilize social conventions, norms or experiences of the subject as firmly rooted and unambiguous. Here, the voices, bodies and dialogue with the listeners in the work help to define, identify and transcend the boundaries and possibilities of the individual presences. In the works *Emigrant* (2013) by Niels Pugholm, Nanna Lysholt Hansens *Miss no. 1, 2, 3* (2013) and *The Woman, the Orphan, and the Tiger* (2010) by Jane Jin Kaisen and Guston Sondin-Kung, there is also a dialogue between work and listener. Anthropologist Amanda Weidman writes in a 2014 *Anthropology and Voice* article that it is in the way in which the voices resonate in listeners and publics that the meaning becomes audible (Weidman 2014: 45). The meaning of these works emerges through the dialogues with the listener. It is dialogues that complicate and question identity categories and belongings, in ways that point to political systems and social issues that exist globally and nationally. The voice as material practice and metaphor is negotiated in the relationship between works and listeners, where the artworks identify and problematize the voice as character, political statements and testimonies. In the following, these works will be discussed with this in mind in an effort to understand what role the voice plays in Danish contemporary art today.

Ill. 1: *Emigrant*, 2013. The white drift clouds moves across the sky as a narrative of emigration unfolds in the work of Niels Pugholm.



### **Emigrant**

"The voice is a sign, just as the smoke from a gun is a sign that it has been fired," says Danish contemporary artist Niels Pugholm (b. 1983) in an interview and continues, "What happens when you speak Danish with a clear accent? What image does it create in the listener? I use the voice as an artistic instrument, a sign value to be interpreted" (Pugholm 2016).

The voice as a sign is one of the elements that Niels Pugholm examines in his work *Emigrant* from 2013 (ill. 1). The work consists of a thirteen-minute-long audio and video installation and has been shown at *Sector 13* at Brandts in 2013 and at *FOCUS 2015* at Nikolaj, Copenhagen Contemporary Art Center.

In the work, we look at drifting clouds that slowly move across a light blue sky and hear a voice say: "The clouds are similar across the globe. It's the same wherever you go. They just hang up there, like little white pillows on a blue sky, or lay over you like a soft blanket. Sometimes they are completely gone, but you are sure they will come back again. I like clouds." The visual imagery is linked to the auditory side of the work. The clouds move quietly from one place to another and out of our field of vision as the work unfolds over time. The voice continues: "I was born in the countryside. My parents had a small farm and we were ten children in the family. I had four sisters who were older than myself, but I was the oldest of the boys." It is a poetic, gentle male voice that speaks Danish with a clear accent. He curls the words softly with rolling r's and waving h's without any punch to them. It indicates that he has a mother tongue other than Danish, though he speaks the language fluently. At this point in the work, we get the impression that the person whose voice we are listening to talks about his own upbringing in the countryside with hard work from an early age. The title, *Emigrant*, and the accent support our coupling of the voice we hear with the story being told. It is an emigration story about leaving one's home country to seek happiness and a better life in another country, in this case Denmark. We understand him well. It is pleasant in Denmark. However, our notion that the emigration story belongs to the voice we listen to crumbles as we hear that the narrative's "I" become ill with the Spanish flu during the great epidemic and a little later, a month after the First World War, leaves for America to try his luck.

The emigration story is inspired by Niels Pugholm's great-grandfather, an ethnic Dane named Christian "Chris" Pedersen (1902-1997). The voice we hear read the story in the work belongs to Georgius Skouros, who emigrated from Greece to Denmark in the 1980s.

### **The voice as a sign**

In Pugholm's work, Skouros' voice is a sign. Just as the smoke from the gun is a sign that it has been fired, is the way Skouros speaks, his fluent Danish with distinct foreign accent, a sign of a special population group who have emigrated from another country to Denmark. At the beginning of the work, we as listeners easily associate Skouros' vocal practice with the story of emigration. We also easily imagine that the work is about a special population group's immigration to Denmark. Thus, in *Emigrant*, Pugholm questions our listening experience of Skouros' vocal practice as a sign of a specific identity, as a metaphor for a person who is not born in Denmark, but who has traveled to seek a better life. Finding that it is not Skouros' own emigration-narrative, our experience of Skouros' voice as a sign of a particular identity is broken. During the work, there is a negotiation between the work and us as listeners, about who it is that we hear speak. The fact that we never get a visual image of the speaker supports the doubts about who really is behind the voice and narrative. The work now appears ambiguous and complex. The voice we listen to is merged with the words, the reading of the narrative in the work, but at the same time each element stands out separately. The voice has been given new meaning as a sign value in the simultaneous merging and separation of the voice as material practice and as a metaphor. The voice we hear links the emigration of the 1910s from Denmark to the United States, with the immigration to Denmark from the 1970s to the present day. The stories are one coherent whole and yet separate. The visual part of the work supports this as the persons are audibly present but not visible physically. In the book, *The Voice in Cinema*, film and sound theorist, Michel Chion, explains the term *acousmètre* as a cinematic tool, an approach by which a voice-character appears without body. The voice wanders on the screen's surface without penetrating it. As a viewer and listener, we can only imagine and fantasize about the body of the voice. In this *acousmètre*, this tension, this equilibrium, wherein we are allowed to hear but not see, there is a force. The voice appears as omnipotent, as ubiquitous (Chion 1999: 24). The voice of *Emigrant* is an *acousmètre*. We never see neither Chris nor Georgius, but instead we follow the clouds drift across the sky, as a global weather phenomenon. The work appears as a poetic conglomerate of the story, the voice and the clouds in the sky. The voice is no longer an identity marker for people who have immigrated to Denmark. The voice is a sign of emigration as a transnational and transhistorical phenomenon. It signifies a personal but also general narrative across different histories and different continents.

Ill. 2: *Miss no. 1, 2, 3* from 2013. In Nanna Lysholt Hansen's work, we are seduced and puzzled by the voices we hear and the words they say.



### Miss no. 1, 2, 3

In 2013, Danish contemporary artist Nanna Lysholt Hansen (b. 1980) created the video work *Miss no. 1, 2, 3* (ill. 2). Among other things, it was shown in 2014 at the exhibition *Body Screeners* in NLHspace, but like many of Lysholt Hansen's other works, *Miss no. 1, 2, 3* are part of a large number of artworks that work with the same subject matter. In 2009, Lysholt Hansen had already created *Miss*, a video work for the Spring Exhibition at Kunsthall Charlottenborg, in which the English word "miss" was read out encircled by words beginning with the "mis-" prefix. *Miss* has also been shown at the exhibition *We Are Grammar* at Pratt Manhattan Gallery in New York in 2011. Later that year, a further development of the work, *Miss no. 1,2,3, live # 1*, in which three glossaries related to the word "miss" were listed as live performance in connection with *Afgang 2011* at the Nikolaj, and later Copenhagen Contemporary Art Centre. The video work *Miss no. 1, 2, 3.*, builds on the previous two works. Here, we hear definitions of "miss" in its meaning as verbs and nouns read side by side with the synonyms of the word, as they are listed in the dictionary. In attempting to identify the word's mechanism of meaning creation and sensuous materiality in the artwork, Lysholt Hansen's work with the voice as an artistic media central. "What kind of voices are we listening to? What does it mean if they are gendered? And what is the difference between whether one or many voices speak?," asks Nanna Lysholt Hansen and continues: "For me it is about asking the questions subtly in the artwork, in a way that opens up the reflection of the listener" (Lysholt Hansen 2016). The question of voice, who or what we listen to, is unfolded in several ways in the work in a sensuous play with the audience's position; who is subject and who is object? Who is being seen and who is watching? What voices are being listened to and how?

In *Miss no. 1, 2, 3* (2013) we see three almost identical video clips of Lysholt Hansen. In all three, we see the artist pictured as a bust, with the same bare shoulders and bright face with hair hanging down in front of the ears. We see them, the clones as Lysholt Hansen calls them, looking directly at us, silent. Then the middle clone starts talking: "Miss, Missed, Missing, Misses". The voice is wooly and deep. It does not sound like a man or a woman, but as something that exists and fluctuates between the sexes. After half a minute, the clone on the left concurs with lists of words with "miss" or "miss-" as a prefix, for example: "mis-apprehend, mis-apprehension, mis-appropriate, mis-appropriation." It doesn't sound like the voice of the clone in the middle. It is a little lighter and lies in the intersection of a tenor and an alto voice. The last clone blends into the word stream: "Miss, a young woman; a young lady of 18, fille, girl, missy, young lady, young woman, babe, sister, baby, sometimes used as a term of address for attractive young women." The voice is like the voices of the other clones, metallic dark. The words attach to the others' dragging, pulsing rhythm. We have doubts as to whether we can place the faces we see within a binary gender classification system in their pairing with the voices. In fragments, the sentences take the form of a



strangely multi-voiced conversation between the three clones. “Miss. Used as a courtesy title before the surname or full name of a girl or single woman,” says the clone in the middle, while the last clone continues with its descriptions: “Bimbo, a young woman indulged by rich and powerful older men.” The words flow in and out of the clones, as if they are the mass-producing word machines. They speak rhythmically, in each other's mouths, next to each other. Faces, voices and words are mixed together; “Used as a form of polite address for a girl or young woman. A young unmarried woman.” Miss 1, 2 and 3 are looking seriously at us with penetrating eyes. We need to look at them. We need to listen to their voices. We need to hear the lists of words they enumerate both together and separately.

### **The voices’ contestation of the politics of words**

Sound theorist Douglas Kahn writes that technology enables a situation and conversation that would not otherwise have occurred – a transformation of materiality, a resolution or change of meaning, in a situation where the artist has set the framework for audience participation in the encounter with the arts (Kahn 1999: 3). Lysholt Hansen works with voice manipulation, timbre and pace, so we are pulled into the rhythm and enticed by the voices' ambiguities and sensual balance between the boundaries of gendered stereotypes. Their tone colors and pitch challenge our experience of the voice as something that can be a metaphor for an identity which are attached to a gender. The work points to how we often equate voice, representation and identity and ignore the complexities of how a voice is actually constructed, mediated and heard. In the work, we hear voices as fundamentally paradoxical: as phenomena that constitute, destabilize and distort our notions of identity. The work shakes our perception of the voice as a metaphor for a subjectivity, as a representation of one speaking subject with one identity and subjectivity. We, as listeners, are not given the opportunity to define, classify and rubricate each single “miss” as a speaking subject with one identity, because they are pitched, reproduced and mediated through different technologies. Neumark writes that the voice is mediated through technologies as well as through cultures (Neumark 2010: xviii). Through a manipulation of the pitch of the voices, this work points to how they are already mediated by our cultural conceptions and how pre-existing understandings of the voice, as something belonging to a gendered body within a binary gender system – the female voice belongs to the female body and the male voice belongs to the male body – dominates our listening of the voice.

Queer theorist and philosopher Judith Butler writes about how naming and denomination constitute a subject through language (Butler 1997: 5). In the book, *Excitable Speech: A Politics of the Performative*, she asks: Can we imagine a body that has not been ascribed to any social definitions? What social existence does this body have? Isn't it first in the calling and addressing that the body comes before us? In the work, indictment and naming are problematized as a way in which language categorizes the body. What happens if the body does not fall into the categories available to us in the language? If one is more; both and, and other than what the indictment covers? Can the words contain us as human beings with a fluid sexuality and gender?

The work's presentation of the social structures that classify voice and body in order for them to be identified within a binary gender system is not only about the voice but also the language. We use the words to address people differently, according to their gender or status. Through the work's destabilization of the voice as a metaphor for an identity, something more also happens. We become strangers to the words we listening to. They have become objects that we can view anew. Why are there different denominations of a woman based on whether or not she is married? And why does this "miss" have synonyms like "bimbo," "sexbomb" and "mistress"? Why are these designations connected to an unmarried woman? Lysholt Hansen's work makes us think critically about the ways

in which language classifies and creates contexts when we speak: for is "miss" and all its associations not a manifestation of political gender-based power structures in the language?

III. 3: *The Woman, The Orphan, and The Tiger* from 2010. As we watch the woman collapse in the courtroom, we hear a cacophonous chorus of voices continuing with a sound of imperative necessity to be heard in the work of Jane Jin Kaisen and Guston Sondin- Kung.



### **The Woman, The Orphan, and The Tiger**

The 72-minute narrative experimental film *The Woman, The Orphan, and The Tiger* from 2010, by the visual artists Jane Jin Kaisen (b. 1980) and Guston Sondin-Kung (b. 1982), creates a strategic political genealogy between three generations of women; the approximately 200,000 women who were subjected to military sexual slavery by the Japanese military between World War I and World War II, the estimated one million women who worked as sex workers around US military bases in South Korea from the 1950s to the present, and the ca. 200,000 children who have been adopted from South Korea by some in the West since the 1950s. Since 2010, the film has been exhibited and screened worldwide, including Videonale13, Germany; Space \*C, Coreana Museum of Art and Jeju Intl. Women's Film Festival, South Korea; Yamagata Intl. Documentary Film Festival, Japan; Taiwan Intl. Documentary Film Festival; DePaul Art Museum, New Wight Gallery, Vox Populi Galleri og Korean American Film Festival New York, USA.

The film uncovers how military and patriarchal violence against women and children became central to geopolitical negotiations between South Korea, the US and Japan, and how this part of the historical record has been extensively silenced and repressed. Yet this history still reverberates in the contemporary moment. This is portrayed in the film through a complex narrative of trauma and memories that are not unfolded in a linear chronology, but instead driven forward in fragmentary sequences that focus on the voices that we hear. The voices are an artistic fulcrum and instrument that is struck from the beginning of the film. Here we see and hear how an elderly

woman behind a witness panel in a courtroom stands up with the assistance of a younger woman to step forward between the witness panel and the judge's chair. The woman's head is bent back and her eyes are almost closed. She turns to the judge and pounding herself with her left hand repeatedly on the heart before she turns back toward the younger woman who reaches out to help her. She declines and turns around towards the courtroom. Her hand moves high over her head. A gathering of younger men and women, probably the jury or lawyers who have remained seated until now, suddenly all leans forward towards the woman as she collapses in the arms of the younger woman.

The context is unclear, but later in the film we find that the excerpt we just saw comes from a war crimes court in Tokyo in 2000, and that the woman has been a sex slave under the control of the Japanese army. We see the whole sequence of the woman's collapse in extreme slow motion. At the same time as the slow motion image settles in us, we hear a voice say: "Trying to figure out how I can talk about something that is unseen [...] that most people don't believe in [...] something really horrible". Another voice joins the first: "I have been having very strange dreams. I had nightmares of being massacred [...]" The first voice emerges again: "The secret or the gap in one's speech gives rise to a ghost". Another voice is introduced: "En følelse som vrede og en følelse som sorg og de følelser hænger sikkert også rigtig meget sammen, det er to følelser som jeg har lagt mærke til..." A fourth voice concurs: "The part we didn't know that was hidden from us and forbidden from us..."

We listen to the meaning of the words, but slowly the words change as the voices accumulate. It becomes difficult to distinguish one voice from the other. Various languages, Danish, English, Korean, Chinese, and accents intersect, amplifying the soundscape so that only fragments of sentences can be heard, and the words dissolve into timbres and rhythms moving back and forth between each other. Fragments of words emerge to disappear back into the polyphonic chorus of intensified voices.

### **The urgent necessity of voices**

As we watch the opening sequence, the elderly woman with her hand pounding on her heart and her body's subsequent collapse, we see how painful memories of past trauma manifest in the body. Her movements demonstrate the inadequacy of the language. As she gives her testimony, her words fall short and the body takes over. The movements express experiences and feelings that have been tormented and suppressed for many years, something that cannot be described adequately with words. Nancy Rose Hunt, a specialist in African history and anthropology, writes about acoustic records as testimonies of the mass rape of women in the Democratic Republic of Congo from 1903 to 2003 (Hunt 2008: 236). She advocates that we not only listen to the words when we listen to testimonies, but that we also listen to *how* the words as materiality contain sounds of shame and sadness. Thus, we can begin to imagine what is not possible to talk about (Hunt 2008: 242). In *The Woman, The Orphan, and The Tiger*, we experience the choir's voices as an audible counterpart to the woman's falling body. The choir becomes a poetic testimony, a way of giving voice to what has been silenced. "In *The Woman, The Orphan, and The Tiger*, there is an emphasis on voices as materiality, tonality. The voices' hesitancy and urgent necessity in being heard," says artist Jane Jin Kaisen (Jin Kaisen 2016). We listen to the choir. The affective sounds and rhythms of the voices that appear and mix with one another. The cacophonous cementing of the voices in the chorus leads us to an alternative listening of the words being said. The semantic listening recedes in the background in favor of an affective listening for the tonality and material intensity of the voices. There is a penetration in the voices that is charged with meaning and relevance.

The multiplicity of voices in the choir should not be seen as a depiction of women's individual personal stories. The choir points beyond them, reaches beyond the individual's testimony and spreads out as a collective. "It is very much because of the content that *The Woman, The Orphan, and The Tiger* have that we chose to start the film with a chorus of voices, so from the beginning there is an emphasis on the collective rather than an individual voice," says Jin Kaisen and continues: "What we wanted with the film was to intervene in the tendency to read these phenomena as isolated and instead look at the underlying structural conditions; the biopolitical mobilization of women who emerged in a nexus of colonialism, militarism, and patriarchy, not just as a Korean problem but as a transnational problem" (Jin Kaisen 2016).

The choir manifests itself as a collective of different voices, thereby materializing the strategic, political kinship between communities in South Korea who have been marginalized and whose stories have been silenced. By marking kinship, a kinship consisting of not being talked about and not being heard – a kinship of ghosts in South Korea, which also reverberates in the diaspora outside the country – the voices point beyond themselves in a contestation of a national discourse of suppression and distance. Through vocal practices, the choir materializes the desire and struggle to officially be regifted as a part of South Korean national history and self-understanding.

In *The Woman, The Orphan, and The Tiger*, the choir creates a double testimony. The choir calls upon us with vibrating intensity. Asking us to listen to the materiality of voices. Clenched hands. A poetic testimony. We, as listeners, become witnesses to what the voices orbit, that which cannot be said. The experiences, conditions and emotions that are rooted in the body; the silencing, the oppression and the alienation. That which one cannot capture in words, that one, instead, has to orbit, but which permeates one's entire presence, *The Tiger*.

### **The dialogue between work and listener**

In *Emigrant, Miss no. 1, 2, 3* and *The Woman, the Orphan, and the Tiger*, three Danish artworks with international dimensions, we are invited to participate in the reflection processes that unfold around the emotional, intimate, bodily, representative and political dimensions of the voice. In the dialogue between works and listeners, voices as metaphor for identity, representation and collectivity, are investigated, constituted, contested and nuanced through poetic staging of vocal practices and expressions. We are involved in the works as active listeners, contributing to the works' meaning production. The works do not provide any clear answers, but instead ask the listener: What is the meaning of voice as representation and marker of identity? What happens when the voice is no longer emitted from a body but, instead, mediated through technologies or systems? And how can one give voice to all who cannot be spoken about?

The voice as an artistic medium lends itself to a form of dialogic meaning production in the work. Weidman writes that the voice is a place where values and understandings, emotions and aesthetic expressions are manifested, debated and challenged through bodily vocal practices (Weidman 2014: 38). We should listen to the works' mobilizations of voices as poetic contestations, investigations and transgressions of identity political issues. As listeners, we are an essential part of the dialogue; it is through our listening to the voices in the work that the work unfolds. Therefore, it is not only about the voices in the artworks, but also how we listen to them; the dialogue between us and the artworks. There is thus an obligation in the relationship. A responsibility to listen to the voices in the artworks; their signs, their political statements, and their testimonies.

### **The paradox of voices**

In VØICE, Neumark writes that the voice was previously perceived as authentic and present, but that now, by virtue of new theories and works, it is possible for us to hear the voice anew, as fundamentally paradoxical, rather than as something that produces, an original presence, identity, and subjectivity (Neumark 2010: xvi). The paradox of the voice is that the voice is one of the most intimate things that can define a person, while never defining the person alone, but always in the world, as an interrelation that occurs between the voice and the listener (LaBelle 2006: 106). The voice transcends the bodily boundaries by which classical subjectivity is defined and lies in the gap. This gap is that of *Emigrant*, *Miss no. 1, 2, 3* and *The Woman, the Orphan, and the Tiger*. The dialogues that unfold between the works and us as listeners enable an experience of the voices we hear as fundamentally paradoxical, ambiguous and complex. By discussing the material vocal practices and how these practices relate to the voice as a metaphor in the works, it has become clear how these artists work with the paradoxicality and complexity of the voice as a medium to articulate identity political issues, immigration, gender and the oppressed voices, without dealing with personal stories or specific people's experiences. Instead, by virtue of the ambiguity of the voices, the voices appear as a contention and criticism of our experience of the voice as authentic, and belonging to an identity and subjectivity. The works destabilize identity categories and put our experience of belonging into a different perspective, such as societal problems that exist on a global scale. These works are internationally oriented and speak to the international contemporary art scene, even when, as in *Emigrant*, they deal with specific Danish conditions. In this way, in dialogue with the listener, they connect questions about identity categories and belonging to issues that exist across continents, languages and populations.



# Technological Fantasies of Nao - Remarks about Alterity Relations

## AUTHOR BIO

**Stina Hasse Jørgensen** is a PhD student at the Department of Arts and Cultural Studies, University of Copenhagen, where she is researching our relationships with machines, specifically in relation to robots. Currently Stina's research concerns the aesthetics of social robots in relation to the conditions of late capitalism. Together with Oliver Alexander Tafdrup from Aarhus University she is conducting experiments with the humanoid robot NAO in a museum context.

Stina has published articles on art, technology and sound in magazines and journals such as *Seismograf*, *Body, Space & Technology Journal* and *Cultural Analysis*, and has given papers at conferences such as *Aesthetics, Ethics and Biopolitics of the Posthuman*, *TERRÆN 2015*, *Historizing the Avant-garde*, *Global Lives Project - UC Berkeley*, *Re-New IMAC*, and *NORDIK2012*.

**Oliver Tafdrup** is a PhD fellow at the Department of Education at Aarhus University, Denmark. His thesis deals with the philosophy of sociotechnical imaginaries. Other research areas include the use of technology in education and the phenomenology of imagination.

Stina Hasse Jørgensen and Oliver Tafdrup

## ABSTRACT

In this contribution we investigate how the concept of 'technological fantasies' can be utilized to further develop the understanding of human-robot relation as an 'alterity relation'. Postphenomenology emphasizes how the humanoid robot is constituted as a 'quasi-other' in the interaction with humans. The basis of the article is an experiment we conducted at the Medical Museion in Copenhagen, involving the humanoid robot Nao as a tour guide. Through interviews with the participants of the robot guided tour, we discuss how technological fantasies of the robot play an active part in the constitution of the alterity relation and thus the experience of the robot as a quasi-other.

## KEYWORDS

Humanoid robots, technological fantasies, sociotechnical imaginaries, postphenomenology, psychoanalysis, critical design

Within Science and Technology Studies (STS), postphenomenology has been the go-to theory when discussing philosophical aspects of human-technology relations, including how we relate to humanoid robots. This phenomenological approach to technology directly addresses how humans relate to robots through an *alterity relation* that establishes the robot as a *quasi-other* (Ihde 102, Coeckelbergh 198). A central aspect of our argument in this article is that the experience of the quasi-otherness of humanoid robots should not just be understood at a phenomenological level; it should also include a discussion of how the design and staging of the robot manipulate and evoke technological fantasies and desires. In the following sections, we will discuss the potential benefits of theorising humanoid robots through a perspective grounded in a combined reading of the postphenomenological conception of alterity relations and the psychoanalytical conception of *fantasy*. The empirical foundation of this article is a *what-if scenario* involving the humanoid robot Nao as a museum guide. The what-if scenario was not designed to optimise the robot's functionalities, but to investigate how we relate to humanoid robots as quasi-others.

### Human-robot Relations: The Staging of the Social

Nao is a 58cm tall humanoid robot developed by Aldebaran Robotics in 2006. Nao has been used in research labs, education, service business and entertainment. Nao is one of the most widely used humanoid robots within academic institutions, where it is often used with the aim of optimising the robot's functionalities or in experiments investigating human-robot interactions. Nao can also be described as a humanoid robot with arms, legs, eyes and fingers as can be seen in the picture of Nao further below. Humanoid robots distinguish themselves from the broader category of sociable robots (Breazeal, *Designing Sociable Robots* 1) by being designed to have humanlike, anthropomorphic features and reactions. The design of humanoid robots aims to promote an experience of robots as having social capacities, i.e. being socially evocative, by encouraging an anthropomorphisation of the robots in order to promote a social interaction with its users (Breazeal, "Toward Sociable Robots" 169). The anthropomorphic aspects are also promoted in the design of humanoid robots; with arms, legs, eyes, and voices that can communicate and interact with their human user and evoke experiences of the robots as social partners. Humanoid robots also tend to be staged, e.g. in commercials and video documentations, as something with social capacities (Suchman 124). This staging of humanoid robots positions them as something subjects can relate to as a quasi-other, or a new companion species for the future human. As the techno-anthropologist Cathrine Hasse writes: "They have been designed to be quasi-others engaging with humans on the road to a fantasy future, where cute sociable robots will engage meaningfully with humans as a new companion species" (Hasse 181). Yet, how can we understand the relation between staging robots as quasi-others and technological fantasies of companion species?



To stage a robot as a social partner, so that it can be related to as an *Other*, is a variant of postphenomenologist Don Ihde's concept of *alterity relations*. Ihde precisely emphasises the robot as an example of what he calls *quasi-otherness* (Ihde 102). The argument that robots are staged quasi-others is thus present in the works of Ihde, but the aspect of how fantasy functions to sustain the experience of otherness, at the level of the subject, is absent. This article provides a contribution to the debate on human-robot relations as alterity relations. In the following, we will investigate how ideologically shaped technological fantasies help to sustain the experience of Nao as a *quasi-other*, even though Nao fails to function properly.

The structure of the article will be tripartite. In the first section, we will elaborate on Ihde's postphenomenological theory and discuss why it is useful in the study of robots. In the second section, we will describe the methodological framework and the setting for our experiment. In the third section, we will apply these theoretical reflections in our analysis of the empirical data acquired through the experiment and discuss how the concept of alterity relations can benefit from psychoanalytical insights into the function of fantasy and desire.

### Postphenomenology and Alterity Relations

Human-technology relations have been a widely debated area within the empirical turn of philosophy of technology (Actheruis 6). The empirical interest in how we relate to technology stems from the break with the classical conception of technology as a specific form of rational force, as we see in e.g. the social philosophy of the critical theorists Theodor Adorno and Max Horkheimer (Adorno & Horkheimer 147) and the phenomenology of the late Martin Heidegger (Heidegger 36). Postphenomenology is a philosophical tradition born out of this turn to empirical analysis of technologies. Rooted in the phenomenological works of Edmund Husserl, the early Martin Heidegger and Maurice Merleau-Ponty, Ihde has developed a phenomenological understanding of how technologies mediate the relation between the embodied consciousness and the world (Ihde 72). In his main work *Technologies and the Lifeworld*, he develops two philosophical programs.

Ihde calls the first program a *Phenomenology of Technics* (Ihde 72). It revolves around the philosophical claim that phenomenology of human-technology relations can be understood through four formalistic structures. The first structure is the *embodiment relation* that draws attention to how technologies can merge with our body and thereby alter our relation to the world; the computers we are using to write this article are an example of this. When typing the words into this document, the keyboard becomes an extension of my fingers. Another everyday example is the bicycle that forms a unity with the body that is riding it. The second structure is the *hermeneutic relation* that draws our attention to how technologies can be read and interpreted like a text. This is the case with the watch through which we read the time. The third structure is the *background relations* which emphasize that most of the

time we find ourselves in complex technical structures, where technologies inconspicuously function in the background. All of these relations can be present in human-robot relations, but the fourth structure is the most interesting to our perspective: the relation between a consciousness and a technological *quasi-other*, which Ihde calls an *alterity relation*. The term *alterity* is rooted in the phenomenology of the French philosopher Emmanuel Levinas, where it designates the radical otherness which an *I* experiences when confronted with an *Other* (Ihde 101). This experience of otherness is, Ihde asserts, also possible when an *I* is confronted with technological artifacts. Robots are of course the example par excellence, but the structure can extend to any kind of technological artifact, in so far as the artifact is in some way anthropomorphized. Alterity relations have, for instance, been discussed by Galit Wellner in relation to our experience of cell phones and Stacey Irwin in connection to our interactions with computers (Irwin). Due to the aim of this article, we will limit our perspectives to humanoid robots. More specifically, how can we understand humanoid robots as a quasi-other? To answer this question we will turn to Ihde's definition of quasi-otherness: "Technological otherness is a *quasi-otherness*, stronger than mere objectness but weaker than the otherness found within the animal kingdom or the human one..." (Ihde 100). This phrase proposes that for objects to be experienced as quasi-others, they must appear to the consciousness as if they are in some way animated. Ihde furthermore states that these animated objects are objects of fascination and they give rise to the experience of interacting with someone (Ibid.). Humanlike characteristics in the appearance of an object can cause a feeling of relating to a quasi-other, but quasi-otherness is specifically apparent in cases where the object has a certain degree of *automation*. The phenomenon of automation seems to enhance the object's animated appearance, and thus the sense that the object has a will of its own or even its own consciousness. This is, arguably, the case with many humanoid robots that are designed for being implemented in social contexts, which were previously reserved for human beings, as Coeckelbergh has also discussed. Before going deeper into the discussion of postphenomenology and human-robot relations, we will briefly introduce Ihde's second program which he calls *cultural hermeneutics* (Ihde 124). Here, Ihde draws attention to the cultural embeddedness of technological artifacts, and we will focus on two philosophical arguments that have become a pivotal aspect of postphenomenology: The claim that 1) artifacts form technological intentionality, and 2) the same technological artifact may be used for different purposes when embedded in different contexts – a phenomenon Ihde calls *multistability* (Ihde 144).

What is technological intentionality? As Ihde writes: "Technologies, by providing a framework for action, do form intentionalities and inclinations within which use-patterns take dominant shape." (Ihde 141). Ihde exemplifies this with the typewriter that, compared to a pen, incites the user to write at a higher speed and to edit more. This argument is relevant to our purposes because it draws attention to how the design of a technological artifact, in our case the robot Nao, incites specific user behaviors. This is related to the concept of multistability, which emphasises that the

technological intentionality and the use-patterns technological artifacts incite are also dependent on the cultural context in which they are performed. Technological artifacts contain the possibility of being stabilized in a variety of use-patterns. We read the purpose, use and possibilities of a technology through our culturally cut lenses (Ihde 146). Ihde's two programs can thus be seen as a micro- and a macro-perspective. The first focuses on the first-person perspective while the latter focuses on how the first-person perspective is always imbedded in a cultural context. These insights have inspired us to view Nao through the perspective of postphenomenology. We see the attempt of introducing a theoretical concept that is sensitive to the function of contemporary ideology as a contribution to the postphenomenological tradition, which does not systematically approach the concept of ideology. In the following sections, we will discuss how the alterity relation of human-robot interaction is affected by technological fantasies. First, we will describe how alterity relations can be explored empirically through making a *what-if* scenario.

### The What-if Scenario

A methodological discussion of how to conduct empirical postphenomenological research into robotics has not been thoroughly unfolded, although the question of a general postphenomenological methodology has been touched upon (Rosenberger & Verbeek 30). We argue that what-if scenarios can provide a methodological framework compliant with postphenomenological mode of analysis and as such function as empirical postphenomenological research. Furthermore, we argue the empirical data from the what-if scenario can elicit aspects of alterity relations through a fictional human-robot interaction. Creating what-if scenarios with humanoid robots can probe critical reflections on the kinds of relations we want to have with robots in the near future, as well as how we experience our relations to humanoid robots today. From this standpoint, we argue that what-if scenario can provide valuable input to research investigating humanoid robots as quasi-others and as technological fantasies.

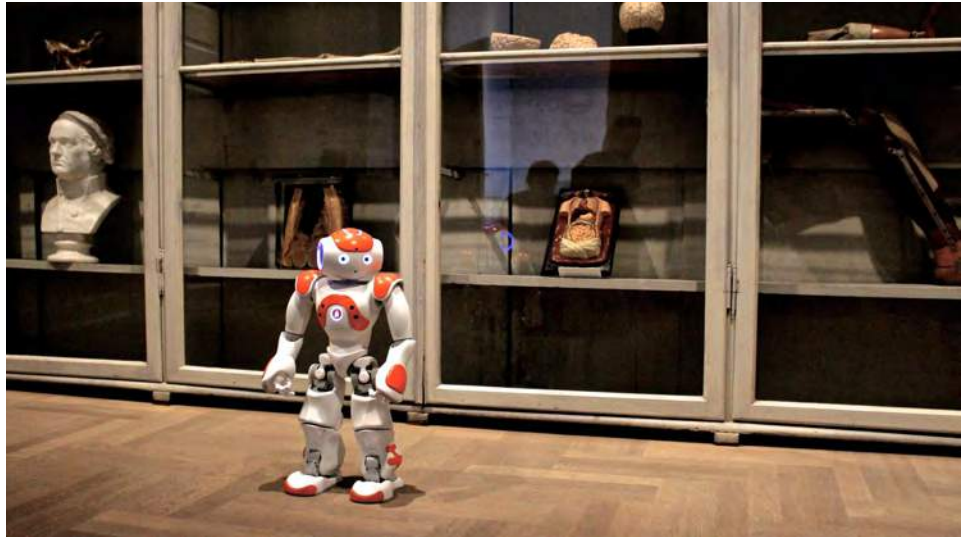
In the following, we will explain how the what-if scenario, as a methodological approach, can be used to investigate the concept of alterity relations in connection to experiences of and interaction with humanoid robots.

The what-if scenario, as it is conceptualised by Fiona Raby and Anthony Dunne, is an approach to investigate alternative uses and relations with existing and emerging technologies and design objects. Dunne and Raby have developed the what-if scenario as an approach in which new technology and design objects are used in fictional situations to stimulate participants in discussions about their near future as well as contemporary reality. Dunne and Raby argue that, as an approach, one can use "design to open up all sorts of possibilities that can be discussed, debated, and used to collectively define

a preferable future for a given group of people: from companies, to cities, to societies” (Dunne & Raby 6).

As part of our research project, we designed a ten minute what-if scenario, which was a robot-guided museum tour that took place on February 11<sup>th</sup> 2016 with twenty four invited participants at the Medical Museion in Copenhagen. The robot-guided museum tour is inspired by Dunne and Raby’s notion of what-if scenarios. In this way the what-if scenario was not designed in order to create better human-robot interactions or investigate how to optimize the robot’s behavior such as other research projects working with robot museum guides have aimed at (Burgard et al.; Pitsch et al.; Thrun et al.).

In collaboration with a guide from the museum, we designed the robot-guided tour of the museum’s display of historical prosthetics. Here, the humanoid robot, Nao, was scripted to perform as a museum guide. The robot-guided museum tour probed the participants to reflect upon the kind of interaction and relation they would have with a museum guide, when the guide was a humanoid robot. In other words, the what-if scenario embodied a near future human-robot relation, and the possible alterity relations that could occur herein, which also functioned as something to be collectively debated in the invited group of participants.



**Fig. 1.** The humanoid robot Nao at Medical Museion.

#### Notes from the Robot-guided Museum Tour:

It took a couple of minutes before the Nao robot’s program was fully loaded, so at first it stood in the corner and looked at the participants. The participants started talking to Nao and waving their hands at it, while waiting for it to start the tour. Many of the participants started filming the robot as it stood quiet in the corner following them with its head. Then it started talking: “Hi, my name is Nao. I will tell you about some of the exhibited objects in this room.” Then the robot

pointed its arm towards the cabinet behind it and started talking about the exhibited prosthetic arms and legs. Before it continued, it asked the participants if they wanted to hear about “the prosthetic arm with the hole in the hand.” The participants echoed each other so the answer: “Yes” became unclear, and Nao’s speech recognition system did not recognize the answer. It paused, bleeped and asked the question again. Then the participants tried answering again, this time more directly towards Nao (with assistance from one of the persons organising the what-if scenario). This time the robot recognized the answer and continued the tour. It told the participants about the different prosthetic objects exhibited in the cabinet. The participants listened to Nao’s descriptions of the design and history of the prosthetic objects and looked at the objects in the cabinet. In the final part of the tour, Nao started to talk about itself as a design object, stating that: “Designers have also spend a lot of time thinking about my aesthetic appearance ... look at my orange shoulders and tiny fingers, I think I am quite cute.” This made the participants laugh. Then Nao asked the participants to touch the person next to them and feel their skin. The participants looked at and touched each other and giggled. Then Nao asked the participants to touch it in order to experience the difference between its robotic arm and the arms of the participants. After some hesitation, one of the participants volunteered and confirmed: “It is different,” which made the group of participants laugh again. Then Nao compared itself to the prosthetics in the cabinet and concluded the tour by saying: “I hope you will think more about prosthetics, robotics as well as the possible future of humans when you are looking at the other exhibited objects exhibited at the Medical Museion.” Most of the participants left the room, but a few of them stayed tickling and touching Nao. It responded: “It tickles.” Then it fell onto the floor and tried to get up. After helping Nao on its feet, the last participants left the room.

**Fig. 2.** Documentation from the robot-guided museum tour at Medical Museion.



### Collective Reflections on Human-robot Relations

A robot-guided museum tour is not yet a reality, but a plausible scenario of what the near future might bring. The what-if scenario was a catalyst for collective action as well as reflections. We designed Nao to interact with the participants in ways that probed them to engage in the robot-guided museum

tour collectively. For example, the participants had to collectively agree on what and how to reply on Nao's questions. In this way, the what-if scenario created an explorative situation where the participants had to reflect and interact with Nao as a group. The what-if scenario created a space for a collective process with unforeseen outcomes and interactions. The participants were co-creators of the what-if scenario. In this way, the what-if scenario had a degree of openness due to the involvement of the group of participants. The collective actions and engagement of the participants were important as we staged the what-if scenario in order to make the participants collectively reflect upon their interaction and experience with Nao as a museum guide. In order to make it possible for the participants to share their reflections on the robot-guided museum tour, we developed an interview guide that the participants could use as a starting point for discussing relevant issues connected to their experiences. In the interview guide, we asked if the participants wanted a robot as a museum guide or what role they wanted the robot to have. We asked how the participants related to the robot; was it as a machine, a quasi-other, an exhibited artifact, an animated presence or something else? The participants debated the questions in pairs of two for around twenty five minutes while audio recording the conversations. They could choose whether to send us the audio recording of their discussion or not. We wanted the participants to feel that they were able to discuss their experiences and relations to Nao critically and without self-censorship. We therefore also made a consent form stating that the participants could decide not to let us quote their discussions, or – if they decided to let us quote their discussions – be anonymised in articles about this project. Ten participants gave us permission to use the reflections and statements that came up during their discussion. This gave us access to interesting perspectives on the human-robot interaction and generated qualitative data, which we analyzed through the prism of critical design, postphenomenology and psychoanalysis. We have chosen not to focus on perceptions of how the robot functions as a museum guide, although this could have been discussed from a museological perspective. Here, our focus is how the participants' establish the robot as a quasi-other and how the staging of the robot's social capacities blends together with the participants' technological fantasies.

### **Alterity Relations and Technological Fantasies**

The what-if scenario created a situation where the participants could observe and interact with a humanoid robot. This situation functioned as a starting point for reflecting upon and discussing how they related to the robot Nao. As such, the what-if scenario framed the discussion of the participants' experience of the humanoid robot.

In the participants' discussion of how they related to the robot, almost all of them focused on the robot's breakdowns. They also focused on the contrasts between their expectations and preconceptions of what the robot could do and how it would interact and what actually happened. We will use the

participants' discussions of how they experienced Nao's breakdowns as well as how the participants experienced Nao as a quasi-other, although Nao failed to function properly.

Here, two of the participating informants, Sara and Kristian, reflect upon the human-robot relation they experienced at the Medical Museion:

Kristian: "Did you stay in the room after the tour was over?"

Sara: "Yes, but I didn't quite understand what happened. I didn't hear it completely"

Kristian: "No, but it was like, you could see that there came much more wonderment in relation to it [Nao]. Like, how you relate to it as a kind of otherness - also in relation to how we talked about it. Because it was sitting and then someone began to stroke the top of its head, and then it said "it's ticklish," and then suddenly it fell onto its back, and said that it needed help. But then "well, okay okay" and then it could even get up on its own - and these things were in fact all the elements that had nothing to do with it being an exhibition robot, but just an otherness, in one way or another.

(Mortensen, Kristian and Larsen, Sara. Dialogue-based interview. 11 February 2016.)

In the above quote, Kristian explains how he, after the tour, experienced the robot as an otherness; what we argue is an alterity relation. The way that it reacted to the tickling and the fact that it was able to get on its feet on its own makes Kristian experience the robot as something more than an object. One could argue that Kristian's experience of the robot as an otherness is a kind of alterity relation, where the robot is a quasi-other to which Kristian relates. This example illustrates that the notion of alterity is a relevant category for understanding human-robot relations. The fact that the informants experience the robot as a quasi-other; i.e. as something more than plastic, screws and scripts, also resonates with two other informants, Louise and Asta:

Asta: That was so nice, he actually became like he reacted with, I don't want to say emotions, but it was like, I felt somehow that there was some kind of consciousness behind that machine, and that was really surprising. I was like, 'is it thinking? Or what is behind this?'

Louise: And when he fell I really felt this too.

(Jensen, Louise and Poulsen, Asta. Dialogue-based interview. 11 February 2016.)

Here, Asta states that the way Nao reacted made her uncertain of whether the humanoid robot has its own consciousness. She describes it as surprising

[1] The concept of *empathy* (Ger. *Einfühlung*) and the related concept of *intersubjectivity* have been much debated within the tradition of phenomenology since Edmund Husserl. See e.g. Zahavi (112). We have chosen not to explore the relation between Ihde's postphenomenology and Zahavi's more cognitive philosophical approach in this article due to our interest in how capitalism shapes our cultural hermeneutic approaches to robots. This interest calls for a theoretical concept that is sensitive to ideology and we propose that the concept of fantasy satisfies this criterion.

to her, as she knows it is a machine. In this way, Asta explains how she experiences Nao as a technological other, as something more than an object, but less than a human or animal. Later in their discussion, Louise and Asta talk more about the otherness they have experienced Nao to be. This indicates that in order to experience the robot as an Other, the participants have to feel a degree of empathy towards the robot. [1] Some of our participants were also relating to Nao as a quasi-other:

Rita: It was almost like I saw his face become like a mask - it comes alive with your imagination. And he blinks and I felt he had facial expressions, even though I know, you know....

Martin: Yeah because he tilted his head, he blinked and so there were some kind of animatedness about him.

Rita: And also the movements, you know, I see him as conscious about himself. His movements; like how to walk and not fall, being careful. I applied human characteristics to him.

Martin: Yeah like on a child level, or an animal in some way. I also thought that he was very humble. Maybe he is programmed to be like that, like "oh, I'm sorry," so that he is not uncanny or frightening. He is just like the most mild, and almost an excuse for himself, when he was trying to get up.

Rita: And also, those mistakes he did, when he didn't understand "yes," so we had to repeat it a second time, I was also just forgiving him that. It was not like I got bored, angry or anything.

(Madsen, Rita and Sander, Martin. Dialogue-based interview. 11 February 2016.)

In the above excerpt, Rita and Martin explain how they experience Nao as animated, but how they also have an experience of interacting *with someone*. They exemplify how they experience Nao as something with personality and characteristics; it is mild, apologetic and careful. Another important part of this discussion is Rita's description of how Nao's face becomes a mask, which she herself animates in her imagination. This can be said to be an example of how Nao is constituted as a fantasy object. It also indicates how her technological fantasy of Nao enables her to maintain an experience of Nao as a quasi-other. Two other participants note:

Laura: You could see that a lot of people tried to get contact with it. But Nao is also built - with the ears, that are actually loud speakers, and the eyes that blink in different colors - to evoke the feelings you have.

Rebekka: Yes, I fall for it completely. Even though I know it is a machine, and that it needs to update as a computer, and that it takes time before it starts up, I want it to be something else.



(Jepsen, Rebekka and Petersen, Laura. Dialogue-based interview. 11 February 2016)

In this quote, Rebekka talks about how she wants to get in contact with the robot, even though she knows she is interacting with a machine. As Laura points out, the design of Nao evokes certain feelings in Rebekka; the way the ears and eyes are blinking creates a wish in Rebekka for the robot to be something else.

Rebekka continues:

Rebekka: I really want to be fooled with an illusion that there is a kind of agency or a presence, lying behind. And I can feel clearly that I am annoyed when it does not live up to my expectations. So something like: “Well, we do not really know just when it starts’ and you could see that there were some technical problems, I can feel that I am annoyed because I would like it to be like a living creature.”

(Jepsen, Rebekka and Petersen, Laura. Dialogue-based interview. 11 February 2016.)

Using the term *fantasy* to account for what is at play in quasi-otherness indicates that the notion of human-robot relations is not exhausted with the post-phenomenological concept of alterity relations. We might very well, at a phenomenological level, experience robotic technology like Nao as quasi-others, but postphenomenology does not tell us much about how technological fantasies are formed, and how they are related to the cultural context, in which Nao is always experienced. We will now take a closer look at the concept of technological fantasies.

### **Technological Fantasies as Ideological Fantasies**

As earlier described, Ihde relates the phenomena of quasi-otherness to a feeling of fascination in the consciousness engaging with the quasi-other. In relation to this discussion, he also uses the term *technological fantasy*, and asserts that there is a “tendency to fantasise its [technological artifacts’] quasi-otherness into an authentic otherness ...” (Ihde 106). Even though Ihde establishes technological fantasy as critical for the experience of technology as quasi-other, he does not elaborate on the concept. In the following, we will argue that the concept of fantasy is crucial for the understanding of how Nao is constituted as a quasi-other, and furthermore how this process is linked to contemporary technological optimism as an aspect of capitalist ideology.

Fantasy is a well-developed theme within Lacanian psychoanalysis. The Slovenian psychoanalytic philosopher, Slavoj Žižek, draws upon the writings of French psychoanalyst Jacques Lacan in order to grasp and critique how

modern capitalist ideologies structure fantasies that subjects can identify with and integrate as a framework for how reality is structured. How does Žižek more specifically understand the term *fantasy*? In his work *The Sublime Object of Ideology*, he asserts an intimate relation between the concept of ideology and the concept of fantasy. According to Žižek, fantasy should not be regarded as something opposed to the objective facts of reality, rather fantasy can be described as a matrix that provides the coordinates of the subject's desire (Žižek 45). On a political level, ideology functions precisely like a fantasy. Hence, Žižek discusses the theme under the heading *ideological fantasy*. The question is now, how do ideological fantasies structure the subject's desire? In Lacanian psychoanalysis, the subject is characterised by a profound ontological lack, while ideological phantasms enable the subject to cope with this lack by constituting commodities, persons, utopias, etc. as objects, which ostensibly can cultivate this lack of being in the subject. The Lacanian/Žižekian notion for these kinds of objects is the *object petit a*. Žižek defines this notion as "the chimerical object of fantasy, the object causing our desire and at the same time - this is its paradox - posed retroactively by this desire" (Žižek 69). An important point related to this is that desired objects are never really what we thought they were when we are confronted with them. As we will argue below, the participants' relation to Nao can be understood through this perspective. How does Nao become "object a"? This point is intimately related to the technological optimism of modern capitalism. Several diagnoses of contemporary capitalism emphasise an optimistic attitude towards technology, where especially information and communication technologies (ICTs) are seen as possessors of a potential for optimising almost any process of knowledge generation (OECD 9; OECD/CERI 17; Boutang 51). The ideological fantasy of technological optimism revolves around a utopian understanding of technology as a cultural force that serves as the driver of economic and cultural improvement (Winner 14). This ideology provides a framework for the subjects to cope with their ontological lack of being, by providing an answer to the question "what am I to the Other?". Desire is closely related to the striving for certainty of one's position in the symbolic order. The fantasy of technological optimism offers the subject a way of dealing with the ontological lack by allowing her to identify as a keen user of technology, capable of seeing the innovative potential of technologies, but, at the same time, it also constitutes technological desire. This is the paradoxical character of desire. As Mladen Dolar writes: "[Fantasy] does not simply procure a phantasmatic object to satisfy desire, it enables the subject to assume any desire at all. There is a strange loop, a circularity of fantasy: it itself fills the lack which it itself opens up and perpetuates" (Dolar xxii). This is the case with technological optimism as an example of contemporary ideology – it offers a way of cultivating the lack of the subject through technology, but at the same time reproduces a lack by constituting technologies as desired object which are not yet attained. Žižek develops this Lacanian conception of fantasy into a critique of ideology that rejects that classical Marxist notion of false consciousness. An important aspect to the Žižekian notion of ideology is that it enables us to sustain our fantasy although, on an intellectual level, we can be critical and distanced. As Žižek writes: "Cynical distance is just one way -

one of many ways - to blind ourselves to the structuring power of ideological fantasy: even if we do not take things seriously, even if we keep an ironical distance, we are still doing them" (Žižek 30).

In an earlier cited quote from one of the research participants, Rebekka states that, even though she knows that Nao is a machine, she wants to experience Nao as something more than a machine. Her comment shows how she keeps a cynical distance; she looks away from what she knows, blinds herself, because she has a desire to experience the robot as something more than a machine.

The ideological illusion is thus to be found on the level of practice, and precisely this point elicits some interesting aspect of alterity relations by providing an answer to why we want to sustain the experience of robots as others, although, on an intellectual level, we know that they are mass produced technological artifacts.

To recapitulate the central points: Our Žižek-inspired discussion of how ideology is related to the concept of technological fantasy provides an opportunity for further developing Ihde's notion of alterity relations. Through Žižek, we have discussed how ideology shapes the technological fantasies through which the robot Nao is constituted as a technological artifact, thus making it appear in an optimistic view, constituting and sustaining the alterity relation between the participants and the robot. Technological fantasies work to enhance and sustain an impression of the robot as a quasi-other, an impression that is also staged and stressed through the design of the robot. The desire to experience the robot as a quasi-other is prevalent in the participants' statements, even though they are aware that the robot does not function properly.

## Conclusion

In this article we have sought to demonstrate how a what-if scenario can be a useful way of conducting empirical research when investigating robotics and postphenomenological questions concerning alterity relations. The what-if scenario created a framework that enabled the participants to reflect upon their technological fantasies and desires in relation to the Nao robot, as well as how this affected their relationship to the robot with which they were interacting. The conflicts between these two levels were in this way framed by the what-if scenario, which enabled nuanced discussions of the participants' technological fantasies and experiences of the Nao robot as a quasi-other or machine, in their interactions with it.

In our analysis of the participants' reflections on the what-if scenario, we have discussed how they related to the robot Nao as a quasi-other and then as "object a." Through a discussion of our empirical data from the what-if scenario, we have attempted to make a more nuanced understanding of the alterity relation by arguing that the Žižekian notion of fantasy can elicit

important elements in our experience of robots as quasi-others, showing how the participants' technological fantasies affect their relation to Nao. Based on the participants' reflections on the what-if scenario, we have discussed how the technological fantasies of the robot, as something more than a machine, lead to a conflict when the desire is not met. We have demonstrated how the participants have experienced the design of Nao as something that stages the robot in a specific way, manipulating technological fantasies and desires emulating humanlike characteristics. This is why the robot appears as something more than a just a mechanical device; viz. a technological otherness or quasi-other. It is not the useful things Nao does during the museum tour that structures the participants' experience of it as a robotic otherness; rather, it is the participants' technological fantasies and desires invested in the robot that structures their experience. In other words, the participants' experience of Nao can be seen as an ideological shaped desire that functions to sustain the alterity relation.

There are obviously some theoretical pitfalls related to the attempt to combine concepts from different theoretical traditions. But despite all the areas where postphenomenology and psychoanalysis are incompatible, the attempt to integrate both the concept of alterity relations and the concept of fantasy has proven fruitful for emphasising that ideology affects the way Nao is experienced as an quasi-other. This discussion is by no means complete with this study. The discussion unfolded above should be seen as a call for further research and debates concerning the connections between humanoid robots, technological fantasies and alterity relations.

## Works Cited

Achterhuis, Hans. "Introduction." *The American Philosophy of Technology - The Empirical Turn*. Ed. Achterhuis, Hans. Bloomington: Indiana University Press, 2001: 1-9.

Adorno, Theodor & Horkheimer, Max. *Dialektik der Aufklärung - Philosophische Fragmente*. Berlin: Fischer Verlage, 2000[1947].

Boutang Yann-Moulier. *Cognitive Capitalism*. Cambridge: Polity, 2011[2008].

Breazeal, Cynthia. *Designing Sociable Robots*. Cambridge, Massachusetts: The MIT Press, 2002.

Breazeal, Cynthia. "Toward Sociable Robots." *Robotics and Autonomous Systems Robotics and Autonomous Systems*, 42.3-4 (2003): 167-175.

Burgard, Wolfram, et. al.. "Experiences with an Interactive Museum Tour-Guide Robot." *Artificial Intelligence (AI)* 114.1-2 (2000): 3-55.

- Coeckelbergh, Mark. "Humans, Animals, and Robots: A Phenomenological Approach to Human-Robot Relations." *International Journal of Social Robotics* 3.2 (2011): 197–204.
- Dolar, Mladen. "Introduction: The Subject Supposed to Enjoy." *Introduction to Grosrichard*. (1998): ix-xxvii.
- Dunne, Anthony and Fiona Raby, *Speculative Everything: Design, Fiction, and Social Dreaming*. Cambridge, Mass.: The MIT Press, 2013.
- Hasse, Cathrine. "Multistable Roboethics." *Ihde Festschrift*. Ed. Berg, Jan Kyrre Friis. New York: Rowan and Atkinson (2015):169-180.
- Heidegger, Martin. *Spørgsmålet om Teknikken*. København: Gyldendal, (1999 [1957/1962]).
- Ihde, Don. *Technology and The Lifeworld*. Bloomington: Indiana University Press.
- Irwin, Stacy. "Technological Other/Quasi Other: Reflection on Lived Experience." *Human Studies* 28.4 (2005): 453-467.
- OECD. *The Knowledge Economy*. Paris, 1996.
- OECD/CERI. *Connected Minds - Technology and Today's Learners*. Paris. 2012.
- Pitsch, Karola, Sebastian Wrede, Jens-Christian Seele, Luise Süßenbach "Attitude of German Visitors Towards an Interactive Art Guide Robot." *Human-Robot Interaction* 6 (2011): 227-228.
- Rosenberger, Robert, & Verbeek, Peter-Paul. *Postphenomenological Investigations: Essays on Human –Technology Relations*, 2015.
- Suchman, Lucy. "Subject Objects." *Feminist Theory* 12.2 (2011): 119–145.
- Thrun, Sebastian et. al.. "Probabilistic Algorithms and the Interactive Museum Tour-Guide Robot Minerva." *Journal of Robotics Research* 19.11(2000): 972-999.
- Wellner, Galit. "The Quasi-Face of the Cell Phone: Rethinking Alterity and Screens." *Human Studies* 37.3(2014): 299–316.
- Winner, Langdon. "Cyberlibertarian Myths and the Prospects for Community." *ACM Sigcas Computers and Society* 27.3 (1997): 14-19.
- Zahavi, Dan. *Self and Other – Exploring Subjectivity, Empathy and Shame*. Oxford: Oxford University Press, 2014.

Žižek, Slavoj. *The Sublime Object of Ideology*. London: Verso, 2008.